STIC Search Report
Biotech-Chem Library

STIC Database Tracking Number: 121495

To: Sarvamangala Devi **Location:** REM 3C18 **Art Unit:** 1645 Thursday, May 13, 2004 **Case Serial Number:** 10/089787	**From:** Beverly Shears **Location:** Remsen Bldg. RM 1A54 **Phone:** 571-272-2528 beverly.shears@uspto.gov

Shears, Beverly

From: Devi, Sarvamangala
Sent: Thursday, May 06, 2004 4:27 PM
To: Shears, Beverly
Subject: 10/089,787

Beverly:

Please perform a sequence and an interference search for SEQ ID NO: 1 and SEQ ID NO: 2 in application SN 10/089,787.

Thanx.

S. DEVI, Ph.D.
AU 1645
Rems - 3C18

STAFF USE ONLY

Date completed: 05-13-04
Searcher: Beverly C 2528
Terminal time: 20
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site
_____ STIC
_____ CM-1
_____ Pre-S

Type of Search
_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors
_____ IG
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
✓ Other CGN

PTO-1590 (9-90)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 08:37:38 ; Search time 46 Seconds
(without alignments)
2469.273 Million cell updates/sec

Title: US-10-089-787-2
Perfect score: 1837
Sequence: 1 MLLLLAEYLQQFYKGFGVFQ.....RFWIITVILVLIGLATLKLR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1689	91.9	360	16	Q88N79	Q88n79 pseudomonas
2	1658	90.3	360	16	Q87WY2	Q87wy2 pseudomonas
3	1314.5	71.6	361	16	Q82VS6	Q82vs6 nitrosomona
4	1306	71.1	360	16	Q8E9P5	Q8e9p5 shewanella
5	1296.5	70.6	361	16	Q83F26	Q83f26 coxiella bu
6	1212.5	66.0	389	16	Q8XVI4	Q8xvi4 ralstonia s
7	1186	64.6	360	16	Q7VP57	Q7vp57 haemophilus
8	1184.5	64.5	377	16	Q7WFR9	Q7wfr9 bordetella
9	1184.5	64.5	377	16	Q7VUQ0	Q7vuq0 bordetella
10	1181.5	64.3	377	16	Q7W4B1	Q7w4b1 bordetella
11	976	53.1	367	16	Q89FU4	Q89fu4 bradyrhizob
12	830.5	45.2	365	16	Q7VQJ0	Q7vqj0 candidatus
13	828	45.1	368	16	Q8KGD1	Q8kgd1 chlorobium
14	819	44.6	368	16	Q8F4J3	Q8f4j3 leptospira
15	782.5	42.6	353	16	Q9PI72	Q9pi72 campylobact
16	704.5	38.4	359	16	Q7VGZ9	Q7vgz9 helicobacte
17	681.5	37.1	326	16	Q8ER51	Q8er51 oceanobacil
18	671	36.5	324	16	Q81WC8	Q81wc8 bacillus an
19	661.5	36.0	318	2	Q8GE12	Q8ge12 heliobacill
20	661	36.0	324	16	Q819Q1	Q819q1 bacillus ce
21	652	35.5	319	16	Q894B9	Q894b9 clostridium
22	639	34.8	321	16	Q8CPK7	Q8cpk7 staphylococ
23	638	34.7	422	16	Q8A255	Q8a255 bacteroides
24	617.5	33.6	357	16	Q82AD9	Q82ad9 streptomyce
25	615.5	33.5	316	16	Q8R9G3	Q8r9g3 thermoanaer
26	611	33.3	321	16	Q88V79	Q88v79 lactobacill
27	572	31.1	368	16	Q8G4Q7	Q8g4q7 bifidobacte
28	555	30.2	373	16	Q7V3S7	Q7v3s7 prochloroco
29	533.5	29.0	369	16	Q83GN4	Q83gn4 tropheryma
30	532.5	29.0	369	16	Q83HJ8	Q83hj8 tropheryma
31	526.5	28.7	339	16	Q8DVM4	Q8dvm4 streptococc
32	524	28.5	359	16	Q7UZF8	Q7uzf8 prochloroco
33	522.5	28.4	378	16	Q8DK95	Q8dk95 synechococc
34	521	28.4	368	16	Q7U3B6	Q7u3b6 synechococc
35	517.5	28.2	370	16	Q7V9F5	Q7v9f5 prochloroco
36	508.5	27.7	336	16	Q8E779	Q8e779 streptococc
37	508.5	27.7	336	16	Q8E1R5	Q8e1r5 streptococc
38	449.5	24.5	312	16	Q9RTD0	Q9rtd0 deinococcus
39	400	21.8	326	10	Q8L7I8	Q8l7i8 arabidopsis
40	284	15.5	356	17	Q8TXC1	Q8txc1 methanopyru
41	254.5	13.9	313	10	Q9M0M0	Q9m0m0 arabidopsis
42	247	13.4	298	17	Q8U3R2	Q8u3r2 pyrococcus
43	246.5	13.4	351	16	Q92D62	Q92d62 listeria in
44	245	13.3	363	16	Q8CX79	Q8cx79 oceanobacil
45	240.5	13.1	357	16	Q81X14	Q81x14 bacillus an

ALIGNMENTS

RESULT 1
Q88N79
ID Q88N79 PRELIMINARY; PRT; 360 AA.
AC Q88N79;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phospho-N-acetylmuramoyl-pentapeptide- transferase.
GN MRAY OR PP1334.
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=160488;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440.";
RL Environ. Microbiol. 4:799-808(2002).
DR EMBL; AE016779; AAN66957.1; -.
DR TIGR; PP1334; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006629; P:lipid metabolism; IEA.
DR GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR InterPro; IPR000715; Glyco_trans_4.
DR InterPro; IPR003524; PNAcPpept_trans.
DR Pfam; PF00953; Glycos_transf_4; 1.
DR PROSITE; PS01347; MRAY_1; 1.

```
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   360 AA;  39307 MW;  91B911C3B4118DE6 CRC64;

  Query Match             91.9%;  Score 1689;  DB 16;  Length 360;
  Best Local Similarity   90.3%;  Pred. No. 1.5e-114;
  Matches  325;  Conservative   18;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||:||| |||||:||||| |||||||:|||||||||||| ||||||||||
Db          1 MLLLLAEYLQQFHKGFAVFQYLSLRGILGVLTALSLALWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||||||| ||||||||||:|||:|||||||||||||||||:||| ||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSAIAVSTLLWADLSNRYVWVVLIVTLAFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||: |||||| || | :|||||:| :| | | ||: |||||
Db        121 IEKNSRGLPSRWKYFWQSVFGLAAAVFLYKTAPTSVETTLILPFIKDVTIPLGVGFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||| |||||||||||||||||||| |||:||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGSGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||| ||:|||||||||||||||||||||||||||||||||||||||||||||:|:|
Db        241 VFCGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGIFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:|||||||||:||||||||||||||||:||||||||||||||||||||||||||
Db        301 TLSVVIQVASFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360


RESULT 2
Q87WY2
ID   Q87WY2      PRELIMINARY;      PRT;   360 AA.
AC   Q87WY2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide- transferase.
GN   MRAY OR PSPTO4411.
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RA   Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA   Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA   Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA   Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA   White O., Fraser C., Collmer A.;
RT   "Complete sequence of Pseudomonas syringae.";
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE016871; AAO57860.1; -.
DR   TIGR; PSPTO4411; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006629; P:lipid metabolism; IEA.
DR   GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   360 AA;  39484 MW;  CF79F93DC0A8CD66 CRC64;

  Query Match             90.3%;  Score 1658;  DB 16;  Length 360;
  Best Local Similarity   89.7%;  Pred. No. 2.6e-112;
  Matches  323;  Conservative   17;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |||||||:|||||||| |||||:||||| |||||:||| ||||||||||: ||||:||||
Db          1 MLLLLAEFLQQFYKGFAVFQYLSLRGILGVLTALTLSLCLGPWMIRTLQMRQIGQSVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||||||| ||||||||||::| |||||||||||||||||:|| ||||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSSIGISTLLWADLSNRYVWVVLLVTFLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||: ||:||| || :  |||||||||| | | ||| |:|||
Db        121 IEKNSRGLPSRWKYFWQSVFGLCAAIFLYTTAPSATETTLIVPMLKDVRIPLGIGFIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||| |||||||||||||||||||| |||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||  ||:|||||||||||||||||||||||||||||||||||||:|||||||:|||||||||
Db        241 VFSGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTMAVIVRQEMVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TLSVVIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360


RESULT 3
Q82VS6
ID   Q82VS6      PRELIMINARY;      PRT;   361 AA.
AC   Q82VS6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MraY; phospho-N-acetylmuramoyl-pentapeptide-transferase (EC
DE   2.7.8.13).
GN   MRAY OR NE0988.
OS   Nitrosomonas europaea.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC   Nitrosomonadaceae; Nitrosomonas.
OX   NCBI_TaxID=915;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19718 / IFO 14298;
RX   MEDLINE=22586410; PubMed=12700255;
RA   Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
RA   Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA   Arciero D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT   "Complete genome sequence of the ammonia-oxidizing bacterium and
RT   obligate chemolithoautotroph Nitrosomonas europaea.";
RL   J. Bacteriol. 185:2759-2773(2003).
DR   EMBL; BX321859; CAD84899.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006629; P:lipid metabolism; IEA.
DR   GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   361 AA;  39264 MW;  AC5BF6C9523B2DD9 CRC64;

  Query Match             71.6%;  Score 1314.5;  DB 16;  Length 361;
  Best Local Similarity   69.8%;  Pred. No. 1.9e-87;
  Matches  252;  Conservative   50;  Mismatches   58;  Indels    1;  Gaps    1;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||| |:::: :  : | || |:||| :|: |||||:|  :|| |||:|   ::||:|||||
Db          1 MLLALSQWIAEDIRAFNVFSYITLRTMLAALTALSISFLIGPAMIRSLTARKVGQSVRND 60
```

```
Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
Db         61 GPQSHLIKAGTPTMGGTLILTAVIVTTLLWADLSNRYIWVVSLTTLGFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIF-FVVL 179
Db        121 IQRNSKGLSASSKFFWQSIIALLVAVYLAMTADLPQHTEMIVPFFKEVAIPLGTFLFIVL 180

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
Db        181 TYLVIVGTSNAVNLTDGLDGLAIMPTVMISGALAIFAYVAGHAVFAKYLGIPHIPNAGEL 240

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
Db        241 AVFCGALTGAGLAFLWFNTYPAEVFMGDVGALALGAALGVITVIVRQEIVLVIMGGVFVM 300

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
Db        301 EALSVMIQVASYKLFGQRVFRMAPLHHHYELKGWKENQVVVRFWIITLILVLIGLSTLKL 360

Qy        360 R 360
Db        361 R 361
```

RESULT 4
Q8E9P5
ID Q8E9P5 PRELIMINARY; PRT; 360 AA.
AC Q8E9P5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phospho-N-acetylmuramoyl-pentapeptide- transferase.
GN MRAY OR SO4222.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=12368813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015855; AAN57194.1; -.
DR TIGR; SO4222; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006629; P:lipid metabolism; IEA.
DR GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR InterPro; IPR000715; Glyco_trans_4.
DR InterPro; IPR003524; PNAcPpept_trans.
DR Pfam; PF00953; Glycos_transf_4; 1.
DR TIGRFAMs; TIGR00445; mraY; 1.
DR PROSITE; PS01347; MRAY_1; 1.
DR PROSITE; PS01348; MRAY_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 360 AA; 39783 MW; CF447768C8F73217 CRC64;

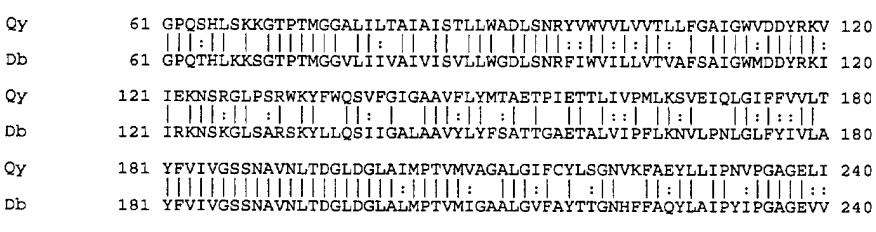
Query Match 71.1%; Score 1306; DB 16; Length 360;
Best Local Similarity 68.6%; Pred. No. 8e-87;
Matches 247; Conservative 47; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
Db          1 MLVYLAEYLTRFHTGFNVFSYVTFRAILGLLTALMFSLWWGPKLIERLQLMQIGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
Db         61 GPESHFSKRGTPTMGGLLILGAIFLSVLLWGDLGSRYVWVMLFVLGSFGMIGFIDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
Db        121 VRKDTKGLIARWKYILQSLAALIIAFFLYTTAANPGETQLVVPFFKDVMPQLGAVFIVLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
Db        181 YFTIVGSSNAVNLTDGLDGLAIMPTVMVAAAFALIAYLSGHAQFANYLHIPHLPGSGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLSLGAALGAIAVLVRQEILLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
Db        301 TVSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISIFLVLLGLATLKLR 360
```

RESULT 5
Q83F26
ID Q83F26 PRELIMINARY; PRT; 361 AA.
AC Q83F26;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phospho-N-acetylmuramoyl-pentapeptide-transferase.
GN MRAY OR CBU0125.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RX MEDLINE=22608657; PubMed=12704232;
RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tettelin H., Davidsen T.M., Beanan M.J.,
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
RT burnetii.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
DR EMBL; AE016960; AAO89689.1; -.
DR TIGR; CBU0125; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006629; P:lipid metabolism; IEA.
DR GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR InterPro; IPR000715; Glyco_trans_4.
DR InterPro; IPR003524; PNAcPpept_trans.
DR Pfam; PF00953; Glycos_transf_4; 1.
DR TIGRFAMs; TIGR00445; mraY; 1.
DR PROSITE; PS01347; MRAY_1; 1.
DR PROSITE; PS01348; MRAY_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 361 AA; 39537 MW; 4B74157AA0729276 CRC64;

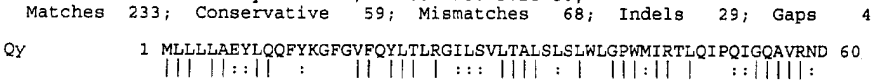
Query Match 70.6%; Score 1296.5; DB 16; Length 361;
Best Local Similarity 67.9%; Pred. No. 3.9e-86;
Matches 245; Conservative 49; Mismatches 66; Indels 1; Gaps 1;

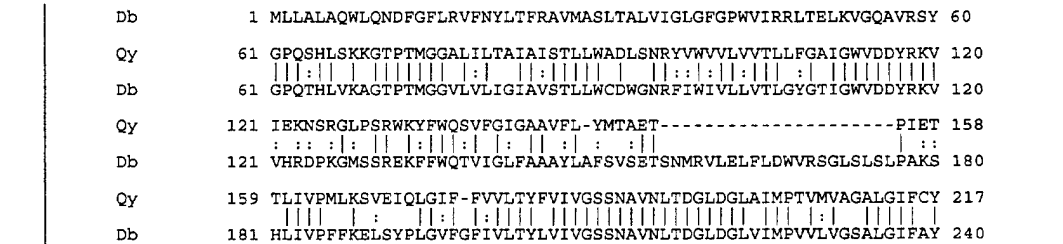
```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
Db          1 MLLWLTNFLSQHFHAFRVFNYLTFRSIVSALTALILVLSLSPRLIKYLVSLQVGQMVRND 60
```



RESULT 6
Q8XVI4
ID Q8XVI4 PRELIMINARY; PRT; 389 AA.
AC Q8XVI4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable HOSPHO-N-acetylmuramoyl-pentapeptide-transferase
DE transmembrane protein (EC 2.7.8.13).
GN MRAY OR RSC2847 OR RS00258.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646072; CAD16554.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006629; P:lipid metabolism; IEA.
DR GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR InterPro; IPR000715; Glyco_trans_4.
DR InterPro; IPR003524; PNAcPpept_trans.
DR Pfam; PF00953; Glycos_transf_4; 1.
DR TIGRFAMs; TIGR00445; mraY; 1.
DR PROSITE; PS01347; MRAY_1; 1.
DR PROSITE; PS01348; MRAY_2; 1.
KW Transferase; Complete proteome.
SQ SEQUENCE 389 AA; 43027 MW; 8897CEAD92997D41 CRC64;

Query Match 66.0%; Score 1212.5; DB 16; Length 389;
Best Local Similarity 59.9%; Pred. No. 5.1e-80;
Matches 233; Conservative 59; Mismatches 68; Indels 29; Gaps 4;





RESULT 7
Q7VP57
ID Q7VP57 PRELIMINARY; PRT; 360 AA.
AC Q7VP57;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phospho-N-acetylmuramoyl-pentapeptide-transferase E.
GN MRAY OR HD0244.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=35000HP / ATCC 700724;
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;
RT "The complete genome sequence of Haemophilus ducreyi.";
RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AE017151; AAP95230.1; -.
KW Transferase; Complete proteome.
SQ SEQUENCE 360 AA; 40242 MW; DD1C6A51AB8C13E7 CRC64;

Query Match 64.6%; Score 1186; DB 16; Length 360;
Best Local Similarity 60.0%; Pred. No. 3.9e-78;
Matches 216; Conservative 61; Mismatches 83; Indels 0; Gaps 0;

```
Db        241 IMCTAIVGAGLGFLWYNTYPAQVFMGDVGSLSLGGALGVIAVLVRQELLLLVMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              |||::|| |:|| :|:|||||||||||||||:||||||||||::|||:|| |||||
Db        301 ALSVILQVGSYKLRQKRIFRMAPIHHHFELKGWPEPRVIVRFWIITLVLVLVGLVTLKLR 360


RESULT 8
Q7WFR9
ID   Q7WFR9      PRELIMINARY;      PRT;   377 AA.
AC   Q7WFR9;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transfera se (EC 2.7.8.13).
GN   MRAY OR MURX OR BB4201.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RB50 / ATCC BAA-588;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640449; CAE34565.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   377 AA;  41315 MW;  3440248BAD69BF1C CRC64;

  Query Match           64.5%;  Score 1184.5;  DB 16;  Length 377;
  Best Local Similarity   62.8%;  Pred. No. 5.3e-78;
  Matches  236;  Conservative   44;  Mismatches   67;  Indels   29;  Gaps    3;

Qy         14 KGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPT 73
              :  |||:|:||| :|:  ||| : |  || :|| |   :|||||  ||:||| | ||||
Db          2 RAIGVFEYITLRAVLACATALLIGLVAGPRVIRRLTEMKIGQAVRAYGPESHLVKTGTPT 61

Qy         74 MGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWK 133
              ||||||| |||||||||| :||:|||||:||  || |||:||||||: ::  |:|:| |
Db         62 MGGALILIAIAISTLLWADWTNRFVWVVLLVTFGFGWIGWMDDYRKVVYRDPEGMPARQK 121

Qy        134 YFWQSVFGIGAAVFLYMTAETPIET----------------------TLIVPMLKSVEIQ 171
              :|||:  |: |||:|     | |                      ||||  |||
Db        122 FFWQATIGLVAAVYLAFAVSAPANTELWPLFKAWVGSGFTMPLPTRADLIVPFFKSVSYP 181

Qy        172 LGIF-FVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLI 230
              ||:  || ||: ||||:||||||||||||||||||||  ||||| |: | | :::|||
Db        182 LGVLGFVALTWAVIVGTSNAVNLTDGLDGLAIMPTVMVGSALGIFAYVVGRVDYSKYLLF 241

Qy        231 PNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVL 290
              | :||| ||:| |||: |||| ||||| ||||||||||||||| ||||||||||||||
Db        242 PYIPGAAELMVLCAAIGGAGLAFLWFNAYPAQVFMGDVGALALGGALGTIAVIVRQEIVL 301

Qy        291 FIMGGVFVMETLSVMIQVASFKLT------GRRVFRMAPIHHHFELKGWPDPRVIVRFWI 344
              ||||||||:||||||:||  || |      |||:|||||:|||||: || : :|:|||||
Db        302 FIMGGVFVVETLSVMVQVTWFKYTKRKYGQGRRIFRMAPLHHHFEVGGWKETQVVVRFWI 361

Qy        345 ITVILVLIGLATLKLR 360
              ||::|||:||:|||||
Db        362 ITMMLVLVGLSTLKLR 377


RESULT 9
Q7VUQ0
ID   Q7VUQ0      PRELIMINARY;      PRT;   377 AA.
AC   Q7VUQ0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transfera se (EC 2.7.8.13).
GN   MRAY OR MURX OR BP3026.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640420; CAE43297.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   377 AA;  41315 MW;  3440248BAD69BF1C CRC64;

  Query Match           64.5%;  Score 1184.5;  DB 16;  Length 377;
  Best Local Similarity   62.8%;  Pred. No. 5.3e-78;
  Matches  236;  Conservative   44;  Mismatches   67;  Indels   29;  Gaps    3;

Qy         14 KGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPT 73
              :  |||:|:||| :|:  ||| : |  || :|| |   :|||||  ||:||| | ||||
Db          2 RAIGVFEYITLRAVLACATALLIGLVAGPRVIRRLTEMKIGQAVRAYGPESHLVKTGTPT 61

Qy         74 MGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWK 133
              ||||||| |||||||||| :||:|||||:||  || |||:||||||: ::  |:|:| |
Db         62 MGGALILIAIAISTLLWADWTNRFVWVVLLVTFGFGWIGWMDDYRKVVYRDPEGMPARQK 121

Qy        134 YFWQSVFGIGAAVFLYMTAETPIET----------------------TLIVPMLKSVEIQ 171
              :|||:  |: |||:|     | |                      ||||  |||
Db        122 FFWQATIGLVAAVYLAFAVSAPANTELWPLFKAWVGSGFTMPLPTRADLIVPFFKSVSYP 181

Qy        172 LGIF-FVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLI 230
              ||:  || ||: ||||:||||||||||||||||||||  ||||| |: | | :::|||
Db        182 LGVLGFVALTWAVIVGTSNAVNLTDGLDGLAIMPTVMVGSALGIFAYVVGRVDYSKYLLF 241

Qy        231 PNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVL 290
              | :||| ||:| |||: |||| ||||| ||||||||||||||| ||||||||||||||
Db        242 PYIPGAAELMVLCAAIGGAGLAFLWFNAYPAQVFMGDVGALALGGALGTIAVIVRQEIVL 301

Qy        291 FIMGGVFVMETLSVMIQVASFKLT------GRRVFRMAPIHHHFELKGWPDPRVIVRFWI 344
              ||||||||:||||||:||  || |      |||:|||||:|||||: || : :|:|||||
Db        302 FIMGGVFVVETLSVMVQVTWFKYTKRKYGQGRRIFRMAPLHHHFEVGGWKETQVVVRFWI 361

Qy        345 ITVILVLIGLATLKLR 360
              ||::|||:||:|||||
Db        362 ITMMLVLVGLSTLKLR 377


RESULT 10
Q7W4B1
ID   Q7W4B1      PRELIMINARY;      PRT;   377 AA.
AC   Q7W4B1;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transfera se (EC 2.7.8.13).
GN   MRAY OR MURX OR BPP3755.
OS   Bordetella parapertussis.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC   Alcaligenaceae; Bordetella.
OX   NCBI_TaxID=519;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=12822 / ATCC BAA-587;
RX   MEDLINE=22827954; PubMed=12910271;
RA   Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA   Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA   Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA   Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA   Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA   Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA   Leather S., Moule S., Norberczak H., O'Neil S., Ormond D., Price C.,
RA   Rabbinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA   Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA   Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT   "Comparative analysis of the genome sequences of Bordetella pertussis,
RT   Bordetella parapertussis and Bordetella bronchiseptica.";
RL   Nat. Genet. 35:32-40(2003).
DR   EMBL; BX640434; CAE39038.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   377 AA;  41327 MW;  63688099A19A7F28 CRC64;

  Query Match             64.3%;  Score 1181.5;  DB 16;  Length 377;
  Best Local Similarity   62.8%;  Pred. No. 8.7e-78;
  Matches  236;  Conservative   43;  Mismatches   68;  Indels   29;  Gaps    3;

Qy         14 KGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPT 73
              :  |||:|:||| :|:  ||| : |  || :|| |  :|||||  ||:||| | ||||
Db          2 RAIGVFEYITLRAVLACATALLIGLVAGPRVIRRLTEMKIGQAVRAYGPESHLVKTGTPT 61

Qy         74 MGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWK 133
              ||||||| |||||||||||  ||:|||||:||  || |||:||||||: ::  |:|:| |
Db         62 MGGALILIAIAISTLLWADWINRFVWVVLLVTFGFGWIGWMDDYRKVVYRDPEGMPARQK 121

Qy        134 YFWQSVFGIGAAVFLYMTAETPIET----------------------TLIVPMLKSVEIQ 171
              :|||:  |: |||:|      | |                          ||||  |||
Db        122 FFWQATIGLVAAVYLAFAVSAPANTELWPLFKAWVGSGFTMPLPTRADLIVPFFKSVSYP 181

Qy        172 LGIF-FVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLI 230
              ||:  || ||: ||||:||||||||||||||||||||  ||||| |: | | :::|||
Db        182 LGVLGFVALTWAVIVGTSNAVNLTDGLDGLAIMPTVMVGSALGIFAYVVGRVDYSKYLLF 241

Qy        231 PNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVL 290
              | :||| ||:| |||: |||| ||||| ||||||||||||||||| |||||||||||||
Db        242 PYIPGAAELMVLCAAIGGAGLAFLWFNAYPAQVFMGDVGALALGGALGTIAVIVRQEIVL 301

Qy        291 FIMGGVFVMETLSVMIQVASFKLT------GRRVFRMAPIHHHFELKGWPDPRVIVRFWI 344
              ||||||||:||||||:||  || |       |||:|||||:|||||: || : :|:|||||
Db        302 FIMGGVFVVETLSVMVQVTWFKYTKRKYGQGRRIFRMAPLHHHFEVGGWKETQVVVRFWI 361

Qy        345 ITVILVLIGLATLKLR 360
              ||::|||:||:|||||
Db        362 ITMMLVLVGLSTLKLR 377


RESULT 11
Q89FU4
ID   Q89FU4      PRELIMINARY;      PRT;   367 AA.
AC   Q89FU4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferas e (EC 2.7.8.3).
GN   MRAY OR BLL6605.
OS   Bradyrhizobium japonicum.
OC   Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC   Bradyrhizobiaceae; Bradyrhizobium.
OX   NCBI_TaxID=375;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=USDA 110;
RX   MEDLINE=22484998; PubMed=12597275;
RA   Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA   Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA   Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA   Tabata S.;
RT   "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT   Bradyrhizobium japonicum USDA110.";
RL   DNA Res. 9:189-197(2002).
DR   EMBL; AP005959; BAC51870.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006629; P:lipid metabolism; IEA.
DR   GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   367 AA;  39570 MW;  A176529AED38E3E4 CRC64;

  Query Match             53.1%;  Score 976;  DB 16;  Length 367;
  Best Local Similarity   52.5%;  Pred. No. 6.5e-63;
  Matches  189;  Conservative   57;  Mismatches  106;  Indels    8;  Gaps    2;

Qy          9 LQQFYKGFG-------VFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQ-IGQAVRND 60
              |   : |||        ||:|:| |   :|:|         |||:|  |:| |  || :| |
Db          8 LSNTFPGFGAFRTFLNVFRYITFRTGGAVVTGALFVFLFGPWIIDHLRIRQGKGQPIRTD 67

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||:||||||||| :||: : :  |:|||:  | |||:|| ||| || :|: ||| ||
Db         68 GPQSHLAKKGTPTMGGLMILSGLTVGTVLWANPLNPYVWIVLAVTLGFGFVGFYDDYLKV 127

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              :: : |  |: :   ::   : |  |       |  | | :| ||   :  | ||||
Db        128 TKQTTTGFGSKLRLLIEAAIALVACYALVRLNRDPASTALTIPFLKDTVLHFGWFFVVFG 187

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||: ||||||||||||:| ::   :  :  ||:||  ||:|| |  | | |||
Db        188 AFVIVGAGNAVNLTDGLDGLAIVPVMIATASFAMIAYLAGNAVFADYLQIKYVAGTGELA 247

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              | | ||:|||||||||  || :|||| |:|||  || ||| |: |||| ::||:||:|
Db        248 VLCGALLGAGLGFLWFNAPPASIFMGDTGSLALGGMLGAIAVAVKHEIVLAVIGGLFVLE 307

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              :||::|| |||||:|:|||||||||  || :|::::|||||:|:| | ||:|||||
Db        308 AVSVIVQVVSFKLTGKRIFRMAPIHHHFEQLGWTEPQIVIRFWIISVMLALAGLSTLKLR 367


RESULT 12
Q7VQJ0
ID   Q7VQJ0      PRELIMINARY;      PRT;   365 AA.
AC   Q7VQJ0;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide transferase (EC 2.7.8.13).
GN   MRAY OR BFL139.
OS   Candidatus Blochmannia floridanus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; ant endosymbionts; Candidatus Blochmannia.
OX   NCBI_TaxID=203907;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22784745; PubMed=12886019;
RA   Gil R., Silva F.J., Zientz E., Delmotte F., Gonzalez-Candelas F.,
```

```
RA   Latorre A., Rausell C., Kamerbeek J., Gadau J., Hoelldobler B.,
RA   van Ham R.C.H.J., Gross R., Moya A.;
RT   "The genome sequence of Blochmannia floridanus: comparative analysis
RT   of reduced genomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:9388-9393(2003).
DR   EMBL; BX248584; CAD83660.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   365 AA;  41504 MW;  E7A2EEFC10966B14 CRC64;

  Query Match             45.2%;  Score 830.5;  DB 16;  Length 365;
  Best Local Similarity   43.4%;  Pred. No. 2.2e-52;
  Matches  155;  Conservative   88;  Mismatches  111;  Indels    3;  Gaps    2;

Qy          1 MLLLLAEYLQQFYKG--FGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIP-QIGQAV 57
              ||  : :    ||    | |   :  ||: :  :| :|: :|  :|  |:   :: | :
Db          1 MLFWIIKITSYFYSSTLFEVMNSVLFRGVGGLFFSLFISIVIGNRIIVWLKYKLRMLQTI 60

Qy         58 RNDGPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDY 117
              | |||||  | |||||| :|| ::  | ::|:|||| |:| :| : :::| :| |||:
Db         61 RIDGPQSHKLKYGTPTMGGIIILISVVTSVIIWSDLSNIYIWYILFIFVMYGILGLVDDF 120

Qy        118 RKVIEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFV 177
              |:   :: ||    || |||:      | :::    :|    :  |:::  :| |: :
Db        121 LKIKRGDNLGLTILNKYLWQSIIAWILIVIMFINRVNYVENQSGLEFLRNIVCKLKIWDM 180

Qy        178 VLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAG 237
              :| ||||||:||:|||:|||||| |:| ::|    | |  :: ||:      | | :|
Db        181 ILAYFVIVGTSNSVNLSDGLDGLVIVPVILVVSGLAIVTWVVGNIYITSDLYIEHVDCIK 240

Qy        238 ELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVF 297
              ||:| ||:::|||||||||:||:|:|||||:|:|  :| ::::: || :| ||||:|
Db        241 ELVVVCASIIGAGLGFLWFNSYPSQIFMGDVGSLSLGGVIGLVSILLHQEYLLLIMGGIF 300

Qy        298 VMETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGL 354
              |:|:|||: ||: ||:  :|:|:||||||||||||: :|:::||||:: ||||: :
Db        301 VIESLSVIFQVSYFKIFKKRIFKMAPIHHHFELKGYMEPKIVVRFWIVSSILVLLSI 357


RESULT 13
Q8KGD1
ID   Q8KGD1      PRELIMINARY;      PRT;   368 AA.
AC   Q8KGD1;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide- transferase.
GN   MRAY OR CT0037.
OS   Chlorobium tepidum.
OC   Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC   Chlorobium.
OX   NCBI_TaxID=1097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TLS / ATCC 49652 / DSM 12025;
RX   MEDLINE=22103685; PubMed=12093901;
RA   Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA   Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA   Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA   Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA   Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA   Vamathevan J., Khouri H., White O., Gruber T.M., Ketchum K.A.,
RA   Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT   "The complete genome sequence of Chlorobium tepidum TLS, a
RT   photosynthetic, anaerobic, green-sulfur bacterium.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR   EMBL; AE012784; AAM71285.1; -.
DR   TIGR; CT0037; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006629; P:lipid metabolism; IEA.
DR   GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   368 AA;  40769 MW;  70F16FA3A3DE4F6A CRC64;

  Query Match             45.1%;  Score 828;  DB 16;  Length 368;
  Best Local Similarity   43.3%;  Pred. No. 3.4e-52;
  Matches  161;  Conservative   78;  Mismatches  117;  Indels   16;  Gaps    5;

Qy          1 MLLLLAEYLQQFYK--GFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVR 58
              ||  :  |: : |   | || :||| |   : :||| : :: |   || |:  :  : ::
Db          1 MLYYILRYINELYSLPGMGVIEYLTFRASAAAITALLIIIFAGQRFIRFLK-SKFVEPIK 59

Qy         59 NDGPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYR 118
               : |  |  ||  ||||| :|: || :|  ||| : : :||:::      | ||::|||:
Db         60 EEAPPEHRKKKDVPTMGGLMIIFAIEVSAFLWAKIDDPHVWLIMLAVFWMGLIGFIDDYQ 119

Qy        119 KVIEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLI----VPMLKSVEIQLGI 174
              ||: |   ||   :|   |   |:  |  : |   |: : |:    ||  | : :  ||
Db        120 KVVLKVKGGLAGHYKLIGQVTLGL---VIGFYTWNDPVFSVLLSDTTVPFFKKLSVDYGI 176

Qy        175 FFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVP 234
              |:: :  |:|   ||||||||||||||      :|  |||:| || ||  :: || || :
Db        177 FYIPVVIFIITAVSNAVNLTDGLDGLAAGNAAIVTFALGVFAYLCGNAVYSGYLSIPFIS 236

Qy        235 GAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMG 294
              ||||: |   |:| | :|||||: ||:|||| |:|:||:|:  ||::::||::| ::
Db        237 GAGEVAVVSMAIVMACVGFLWFNSSPAEVFMGDTGSLSLGSAIAVIALMIKQELLLPVLA 296

Qy        295 GVFVMETLSVMIQVASFKLT------GRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVI 348
              ||| :|||| :||| ||::      |||:| |||:||||:|||| : ::::|||||::
Db        297 GVFFVETLSVSMQVAWFKISKKLYGEGRRIFLMAPLHHHFQLKGWAEQKIVIRFWIISIL 356

Qy        349 LVLIGLATLKLR 360
              | |  | |||||
Db        357 LFLTSLMTLKLR 368


RESULT 14
Q8F4J3
ID   Q8F4J3      PRELIMINARY;      PRT;   368 AA.
AC   Q8F4J3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide- transferase (EC 2.7.8.13).
GN   MRAY1 OR LA2048.
OS   Leptospira interrogans.
OC   Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX   NCBI_TaxID=173;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA   Ren S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE011377; AAN49247.1; -.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR   GO; GO:0016740; F:transferase activity; IEA.
DR   GO; GO:0006629; P:lipid metabolism; IEA.
DR   GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   368 AA;  40644 MW;  03071CAD6DE50C80 CRC64;

  Query Match             44.6%;  Score 819;  DB 16;  Length 368;
```

```
  Best Local Similarity   45.8%;  Pred. No. 1.5e-51;
  Matches  167;  Conservative   66;  Mismatches  120;  Indels   12;  Gaps    3;

Qy          8 YLQQFY----KGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQ 63
              ||    |        :| |:| | ::: ||:: :: | |  :|  |   :  ::||:|||:
Db          4 YLYDLYFNHLDSLRIFSYVTFRALMAGLTSMLVTFWFGHKIIDFLYGLKFRESVRDDGPK 63

Qy         64 SHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEK 123
              ||  |||||||||| ||: :: || ||| :| |  | :: | :| |  :|: ||| | ::|
Db         64 SHEIKKGTPTMGGLLIIGSLLISVLLWGNLKNPNVILLSVFSLSFSVLGFADDYMKSVKK 123

Qy        124 NSRGLPSRWKYFWQSVFG-IGAAVFLYMTAETPIE-------TTLIVPMLKSVEIQLGIF 175
                |: :| |:    :   |  :| | | |           | |  | :|   | |||
Db        124 IKGGMRARTKFILSILISFIFCILFFYYTGTTGQTGKISFQLTDLFFPFIKGPVIALGII 183

Qy        176 FVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPG 235
               :  :  ||:|||:||||||||||||||    ::    ||:  | ||    | || || :||
Db        184 AIPFSILVIIGSSHAVNLTDGLDGLATGTVLISVMTLGVIAYFSGTPIVANYLNIPYLPG 243

Qy        236 AGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGG 295
              | |  || :|| ||  |||||| :|||||||| |:| ||| || | ::::::||:| |:|
Db        244 AHEYSVFLSALTGALFGFLWFNAHPAQVFMGDTGSLFLGATLGMIVILLKKEILLLILGA 303

Qy        296 VFVMETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLA 355
              :|| | |||::|| ||||||:|:|:|||:|||||| |  : ::::|||| ||| :| |:
Db        304 IFVSEALSVILQVGSFKLTGKRIFKMAPLHHHFELGGLKETKIVIRFWIIAVILAIISLS 363

Qy        356 TLKLR 360
              |||::
Db        364 TLKIQ 368


RESULT 15
Q9PI72
ID   Q9PI72      PRELIMINARY;      PRT;   353 AA.
AC   Q9PI72;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide- transferase (EC 2.7.8.13).
GN   MRAY OR CJ0433C.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC   Campylobacteraceae; Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
DR   EMBL; AL139075; CAB74269.1; -.
DR   PIR; G81387; G81387.
DR   GO; GO:0016020; C:membrane; IEA.
DR   GO; GO:0003824; F:catalytic activity; IEA.
DR   GO; GO:0008963; F:phospho-N-acetylmuramoyl-pentapeptide-trans. . .; IEA.
DR   GO; GO:0006629; P:lipid metabolism; IEA.
DR   GO; GO:0009252; P:peptidoglycan biosynthesis; IEA.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Complete proteome.
SQ   SEQUENCE   353 AA;  39239 MW;  8EBBCFC4F1061C95 CRC64;

  Query Match           42.6%;  Score 782.5;  DB 16;  Length 353;
  Best Local Similarity   44.8%;  Pred. No. 6.4e-49;
  Matches  155;  Conservative   70;  Mismatches  118;  Indels    3;  Gaps    2;

Qy         16 FGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPTMG 75
              :  | |:::|   :   || |||:| |  |  :     | :    |::| :|  |||||
Db         10 YAFFTYISVRAGFAFFIALCLSLFLMPKFITWAKAKNASQPIYEYAPETHKTKCHTPTMG 69

Qy         76 GALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEK-NSRGLPSRWKY 134
              | : :::  |::|       | :    |:  :||  || :||  ||::| |  ||  | |
Db         70 GLIFISSAVIASLFCIKFDNIFAISALLCLILFCLIGLIDDLGKVLKKDNHSGLSPRMKL 129

Qy        135 FWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLT 194
              | : |:   : || ::|  : | | :|  |     : || :     |:: ||||||||
Db        130 LAQIIAGLICILPLYFSSE--LSTELFIPFYKHPLFDMEIFAIAFWILVLISSSNAVNLT 187

Qy        195 DGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFL 254
              ||||||| :| :      |||| |||||: ::||||:| : | ||::: ||||:|| :|||
Db        188 DGLDGLATVPGIFSLSTLGIFLYLSGNLNYSEYLLLPKIQGLGEVVIICAALIGALMGFL 247

Qy        255 WFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVMETLSVMIQVASFKLT 314
              |:| |||||||| |:|||  :| :||| : ||:| ::| |||:||:||::|| |||:
Db        248 WYNCYPAQVFMGDSGSLALGGFIGFLAVISKNEILLLLIGFVFVLETVSVILQVGSFKIF 307

Qy        315 GRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
               :|||:|||||||||  || : ::|||||:| ::  |: ||::|||
Db        308 NKRVFKMAPIHHHFEKVGWVENKIIVRFWMIALLSNLLALASIKLR 353


Search completed: May  7, 2004, 08:40:32
Job time : 58 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 08:37:37 ; Search time 18 Seconds
(without alignments)
1041.403 Million cell updates/sec

Title: US-10-089-787-2
Perfect score: 1837
Sequence: 1 MLLLLAEYLQQFYKGFGVFQ.....RFWIITVILVLIGLATLKLR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1824	99.3	360	1	MRAY_PSEAE	Q9hvz8 pseudomonas
2	1689	91.9	360	1	MRAY_PSEPK	Q88n79 pseudomonas
3	1658	90.3	360	1	MRAY_PSESM	Q87wy2 pseudomonas
4	1331.5	72.5	361	1	MRAY_XANAC	Q8ppb0 xanthomonas
5	1327.5	72.3	361	1	MRAY_XANCP	Q8pck2 xanthomonas
6	1312	71.4	360	1	MRAY_SHEVI	Q9f1n3 shewanella
7	1306	71.1	360	1	MRAY_SHEON	Q8e9p5 shewanella
8	1296.5	70.6	361	1	MRAY_XYLFA	Q9pf83 xylella fas
9	1293	70.4	360	1	MRAY_VIBPA	Q87sg7 vibrio para
10	1292.5	70.4	361	1	MRAY_XYLFT	Q87af7 xylella fas
11	1292	70.3	360	1	MRAY_VIBCH	Q9kpg4 vibrio chol
12	1291	70.3	360	1	MRAY_VIBVU	Q8dek7 vibrio vuln
13	1274	69.4	360	1	MRAY_PASMU	P57816 pasteurella
14	1268	69.0	360	1	MRAY_ECOL6	Q8x9z0 escherichia
15	1268	69.0	360	1	MRAY_ECOLI	P15876 escherichia
16	1265	68.9	360	1	MRAY_YERPE	Q8zif2 yersinia pe
17	1264	68.8	360	1	MRAY_SALTI	Q8z9h1 salmonella
18	1261	68.6	360	1	MRAY_SALTY	Q8zru5 salmonella
19	1216	66.2	360	1	MRAY_HAEIN	P45062 haemophilus
20	1186	64.6	360	1	MRAY_HAEDU	Q7vp57 haemophilus
21	1135	61.8	360	1	MRAY_WIGBR	Q8d2z3 wiggleswort
22	1100	59.9	360	1	MRAY_NEIMB	Q9k0y6 neisseria m
23	1099	59.8	360	1	MRAY_NEIMA	Q9jsz3 neisseria m
24	945	51.4	360	1	MRAY_BRUME	Q8yi69 brucella me
25	939	51.1	360	1	MRAY_RHILO	Q98kb0 rhizobium l
26	931	50.7	366	1	MRAY_RHIME	Q52952 rhizobium m
27	917	49.9	366	1	MRAY_AGRT5	Q8udm5 agrobacteri
28	915.5	49.8	357	1	MRAY_BUCAI	P57314 buchnera ap
29	898.5	48.9	361	1	MRAY_RICMO	Q9akp2 rickettsia
30	894	48.7	356	1	MRAY_ZYMMO	P56834 zymomonas m
31	893.5	48.6	361	1	MRAY_RICRI	Q9aki9 rickettsia
32	890.5	48.5	361	1	MRAY_RICCN	Q92h61 rickettsia
33	888.5	48.4	361	1	MRAY_RICTY	Q9akd8 rickettsia
34	887.5	48.3	361	1	MRAY_RICPR	Q9zcw0 rickettsia
35	875	47.6	340	1	MRAY_BUCBP	P59436 buchnera ap
36	853.5	46.5	359	1	MRAY_AQUAE	O66465 aquifex aeo
37	819	44.6	361	1	MRAY_FUSNN	Q8rdq0 fusobacteri
38	812.5	44.2	351	1	MRAY_BORBU	Q44776 borrelia bu
39	709.5	38.6	353	1	MRAY_HELPJ	Q9zly1 helicobacte
40	708.5	38.6	353	1	MRAY_HELPY	O25235 helicobacte
41	671	36.5	325	1	MRAY_BACHD	Q9k9s6 bacillus ha
42	652	35.5	321	1	MRAY_STAAM	O07322 staphylococ
43	652	35.5	321	1	MRAY_STAAW	Q8nx36 staphylococ
44	651	35.4	324	1	MRAY_BACSU	Q03521 bacillus su
45	639	34.8	321	1	MRAY_STAEP	Q8cpk7 staphylococ

ALIGNMENTS

```
RESULT 1
MRAY_PSEAE
ID   MRAY_PSEAE     STANDARD;      PRT;   360 AA.
AC   Q9HVZ8; Q9EY47;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR PA4415.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21178826; PubMed=11281713;
RA   Azzolina B.A., Yuan X., Anderson M.S., El-Sherbeini M.;
RT   "The cell wall and cell division gene cluster in the Mra operon of
RT   Pseudomonas aeruginosa: cloning, production, and purification of
RT   active enzymes.";
RL   Protein Expr. Purif. 21:393-400(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AY008276; AAG22121.1; -.
DR   EMBL; AE004856; AAG07803.1; -.
DR   PIR; H83094; H83094.
```

```
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     21     43       POTENTIAL.
FT   TRANSMEM     71     90       POTENTIAL.
FT   TRANSMEM     97    114       POTENTIAL.
FT   TRANSMEM    134    156       POTENTIAL.
FT   TRANSMEM    168    185       POTENTIAL.
FT   TRANSMEM    200    222       POTENTIAL.
FT   TRANSMEM    235    257       POTENTIAL.
FT   TRANSMEM    262    284       POTENTIAL.
FT   TRANSMEM    289    311       POTENTIAL.
FT   TRANSMEM    338    357       POTENTIAL.
FT   CONFLICT     51     51       R -> P (IN REF. 1).
FT   CONFLICT    335    335       E -> D (IN REF. 1).
SQ   SEQUENCE   360 AA;  39653 MW;  DE8BEE7557091CB3 CRC64;

  Query Match             99.3%;  Score 1824;  DB 1;  Length 360;
  Best Local Similarity   99.4%;  Pred. No. 2.1e-122;
  Matches  358;  Conservative    1;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360


RESULT 2
MRAY_PSEPK
ID   MRAY_PSEPK     STANDARD;      PRT;   360 AA.
AC   Q88N79;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR PP1334.
OS   Pseudomonas putida (strain KT2440).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=160488;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22423060; PubMed=12534463;
RA   Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA   Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA   Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J.,
RA   Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I.,
RA   Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A.,
RA   Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H.,
RA   Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA   Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA   Fraser C.M.;
RT   "Complete genome sequence and comparative analysis of the
RT   metabolically versatile Pseudomonas putida KT2440.";
RL   Environ. Microbiol. 4:799-808(2002).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE016779; AAN66957.1; -.
DR   TIGR; PP1334; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     21     43       Potential.
FT   TRANSMEM     71     90       Potential.
FT   TRANSMEM     97    114       Potential.
FT   TRANSMEM    134    156       Potential.
FT   TRANSMEM    168    185       Potential.
FT   TRANSMEM    200    222       Potential.
FT   TRANSMEM    229    251       Potential.
FT   TRANSMEM    261    283       Potential.
FT   TRANSMEM    288    310       Potential.
FT   TRANSMEM    338    357       Potential.
SQ   SEQUENCE   360 AA;  39307 MW;  91B911C3B4118DE6 CRC64;

  Query Match             91.9%;  Score 1689;  DB 1;  Length 360;
  Best Local Similarity   90.3%;  Pred. No. 7.6e-113;
  Matches  325;  Conservative   18;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |||||||||||:||| ||||||:|||| |||||||:|||||||||||| |||||||||
Db          1 MLLLLAEYLQQFHKGFAVFQYLSLRGILGVLTALSLALWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||| ||||||||||:|||:|||||||||||||||:||| ||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSAIAVSTLLWADLSNRYVWVVLIVTLAFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              |||||||||||||||||||: |||||| || | :|||||:| :| | | ||: |||||
Db        121 IEKNSRGLPSRWKYFWQSVFGLAAAVFLYKTAPTSVETTLILPFIKDVTIPLGVGFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||||| |||||||||||||||||||| |||:||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGSGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||| ||:|||||||||||||||||||||||||||||||||||||||||||||:||:|
Db        241 VFCGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGIFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:|||||||||:|||||||||||||||||:|||||||||||||||||||||||||||
```

```
Db        301 TLSVVIQVASFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360

RESULT 3
MRAY_PSESM
ID   MRAY_PSESM     STANDARD;      PRT;   360 AA.
AC   Q87WY2;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR PSPTO4411.
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=22834015; PubMed=12928499;
RA   Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA   Gwinn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F.,
RA   Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H.,
RA   Nelson W.C., Davidsen T., Zafar N., Zhou L., Liu J., Yuan Q.,
RA   Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T.,
RA   Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.-L., Ramos A.R.,
RA   Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P.,
RA   Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L.,
RA   White O., Fraser C.M., Collmer A.;
RT   "The complete genome sequence of the Arabidopsis and tomato pathogen
RT   Pseudomonas syringae pv. tomato DC3000.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE016871; AAO57860.1; -.
DR   TIGR; PSPTO4411; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM      21     43       Potential.
FT   TRANSMEM      71     90       Potential.
FT   TRANSMEM      97    114       Potential.
FT   TRANSMEM     134    156       Potential.
FT   TRANSMEM     168    185       Potential.
FT   TRANSMEM     200    222       Potential.
FT   TRANSMEM     235    257       Potential.
FT   TRANSMEM     262    284       Potential.
FT   TRANSMEM     289    311       Potential.
FT   TRANSMEM     338    357       Potential.
SQ   SEQUENCE   360 AA;  39484 MW;  CF79F93DC0A8CD66 CRC64;

  Query Match             90.3%;  Score 1658;  DB 1;  Length 360;
  Best Local Similarity   89.7%;  Pred. No. 1.2e-110;
  Matches  323;  Conservative   17;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |||||||:|||||||| ||||||:||||| ||||||:||| |||||||||||: |||||:||||
Db          1 MLLLLAEFLQQFYKGFAVFQYLSLRGILGVLTALTLSLCLGPWMIRTLQMRQIGQSVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||||||| |||||||||||::| |||||||||||||||||:|| |||||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSSIGISTLLWADLSNRYVWVVLLVTFLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              |||||||||||||||||||||: ||:||| || :  |||||||||| | | ||| |:|||
Db        121 IEKNSRGLPSRWKYFWQSVFGLCAAIFLYTTAPSATETTLIVPMLKDVRIPLGIGFIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||| ||||||||||||||||||||| ||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||  ||:|||||||||||||||||||||||||||||||||||:|||||||:||||||||||
Db        241 VFSGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTMAVIVRQEMVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:|||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        301 TLSVVIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360


RESULT 4
MRAY_XANAC
ID   MRAY_XANAC     STANDARD;      PRT;   361 AA.
AC   Q8PPB0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR XAC0777.
OS   Xanthomonas axonopodis (pv. citri).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=92829;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=306 / ATCC 13902 / XV 101;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE011708; AAM35665.1; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     21     43       POTENTIAL.
FT   TRANSMEM     71     90       POTENTIAL.
FT   TRANSMEM     97    114       POTENTIAL.
FT   TRANSMEM    134    152       POTENTIAL.
FT   TRANSMEM    165    187       POTENTIAL.
FT   TRANSMEM    202    224       POTENTIAL.
FT   TRANSMEM    231    253       POTENTIAL.
FT   TRANSMEM    263    285       POTENTIAL.
FT   TRANSMEM    290    312       POTENTIAL.
FT   TRANSMEM    339    358       POTENTIAL.
SQ   SEQUENCE   361 AA;  39626 MW;  BC44B2A5027348AB CRC64;

  Query Match             72.5%;  Score 1331.5;  DB 1;  Length 361;
  Best Local Similarity   70.4%;  Pred. No. 1.6e-87;
  Matches  254;  Conservative   43;  Mismatches   63;  Indels    1;  Gaps    1;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||| || :|||    ||:| ||| |||| |||| ||||:|| :|| |   : || :| |
Db          1 MLLELARWLQQLESLFGLFNYLTFRGILAALTALFLSLWMGPAVIRKLAQFKGGQPIRQD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||:| || |||||||:|||  : :| ||| || |||||:|| | : |||||| ||: |:
Db         61 GPQTHFSKAGTPTMGGSLILLTVTLSVLLWGDLRNRYVWLVLAVMICFGAIGWYDDWIKI 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQL-GIFFVVL 179
              ::::  || |||||  ||:||: | :||| ||: |   |  :|| ||: : | |: |||:
Db        121 VKRDPNGLKSRWKYLLQSIFGLAAGLFLYYTADVPAAITFYIPMFKSIALPLAGVSFVVI 180

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
              || ||| |||||||||||||||||:|| |||:| | |||| ||||| || :||||||
Db        181 AYFWIVGFSNAVNLTDGLDGLAIMPTVLVACALGVFAYASGNVVFAEYLKIPLIPGAGEL 240

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
              |: |:|: |||||||||||||| |||||:|||:||| |||||||||||:|| |||||||:
Db        241 IIICSAIAGAGLGFLWFNTYPAMVFMGDIGALSLGAVLGTIAVIVRQEMVLVIMGGVFVI 300

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              |||||||||||||||:|||||||||||||||||:|||||||||:|:|||||||||:
Db        301 ETLSVMIQVASFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIISVVLVLIGLATLKV 360

Qy        360 R 360
              |
Db        361 R 361


RESULT 5
MRAY_XANCP
ID   MRAY_XANCP     STANDARD;      PRT;   361 AA.
AC   Q8PCK2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR XCC0723.
OS   Xanthomonas campestris (pv. campestris).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xanthomonas.
OX   NCBI_TaxID=340;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33913 / NCPPB 528;
RX   MEDLINE=22022145; PubMed=12024217;
RA   da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA   Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA   Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA   Camarotte G., Cannavan F., Cardozo J., Chambergo F., Ciapina L.P.,
RA   Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA   Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA   Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA   Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA   Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA   Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA   Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA   Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA   Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA   Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA   Setubal J.C., Kitajima J.P.;
RT   "Comparison of the genomes of two Xanthomonas pathogens with differing
RT   host specificities.";
RL   Nature 417:459-463(2002).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE012171; AAM40038.1; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     21     43       POTENTIAL.
FT   TRANSMEM     71     90       POTENTIAL.
FT   TRANSMEM     97    114       POTENTIAL.
FT   TRANSMEM    134    152       POTENTIAL.
FT   TRANSMEM    165    187       POTENTIAL.
FT   TRANSMEM    202    224       POTENTIAL.
FT   TRANSMEM    231    253       POTENTIAL.
FT   TRANSMEM    263    285       POTENTIAL.
FT   TRANSMEM    290    312       POTENTIAL.
FT   TRANSMEM    339    358       POTENTIAL.
SQ   SEQUENCE   361 AA;  39606 MW;  FF547C44E4FD21B8 CRC64;

  Query Match             72.3%;  Score 1327.5;  DB 1;  Length 361;
  Best Local Similarity   69.8%;  Pred. No. 3.1e-87;
  Matches  252;  Conservative   44;  Mismatches   64;  Indels    1;  Gaps    1;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||| || :|||    ||:| ||| |||| |||| ||||:|| :|| |   : || :| |
```

```
Db          1 MLLELARWLQQLESLFGLFNYLTFRGILAALTALFLSLWMGPAVIRKLAQFKGGQPIRQD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||:| || |||||||:|||  : :| ||| || |||||:|| | : |||||| ||: |:
Db         61 GPQTHFSKAGTPTMGGSLILLTVTLSVLLWGDLRNRYVWLVLAVMICFGAIGWYDDWIKI 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQL-GIFFVVL 179
              : ::  || |||||  ||:||: | :||| ||: | |  :|| |:: : | |: |||:
Db        121 VRRDPNGLKSRWKYLLQSIFGLAAGLFLYYTADVPAAITFYIPMFKAIALPLAGVSFVVI 180

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
               || ||| |||||||||||||||||:|| |||:| | |||| ||||| || :||||||
Db        181 AYFWIVGFSNAVNLTDGLDGLAIMPTVLVACALGVFAYASGNVVFAEYLKIPLIPGAGEL 240

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
              |: |:|: |||||||||||||| |||||:|||:||| |||:|||||||:|| ||||||:
Db        241 IIICSAIAGAGLGFLWFNTYPAMVFMGDIGALSLGAVLGTVAVIVRQELVLVIMGGVFVI 300

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              |||||||||||||||:|||||||||||||||||:||||||||:|:||||||||||:
Db        301 ETLSVMIQVASFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIISVVLVLIGLATLKV 360

Qy        360 R 360
              |
Db        361 R 361
```

```
RESULT 6
MRAY_SHEVI
ID   MRAY_SHEVI     STANDARD;      PRT;   360 AA.
AC   Q9F1N3;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY.
OS   Shewanella violacea.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC   Alteromonadaceae; Shewanella.
OX   NCBI_TaxID=60217;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ishii A., Nakasone K., Sato T., Sugai M., Wachi M., Nagai K., Kato C.;
RT   "Isolation and characterization of dcw gene cluster for cell division
RT   and cell wall synthesis from a deep-sea piezophilic Shewanella
RT   violacea.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB052554; BAB19199.1; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane.
FT   TRANSMEM     20     42       POTENTIAL.
FT   TRANSMEM     74     93       POTENTIAL.
FT   TRANSMEM     97    114       POTENTIAL.
FT   TRANSMEM    134    156       POTENTIAL.
FT   TRANSMEM    171    193       POTENTIAL.
FT   TRANSMEM    200    219       POTENTIAL.
FT   TRANSMEM    234    256       POTENTIAL.
FT   TRANSMEM    263    285       POTENTIAL.
FT   TRANSMEM    289    311       POTENTIAL.
FT   TRANSMEM    338    357       POTENTIAL.
SQ   SEQUENCE   360 AA;  39777 MW;  53B2A2BAEF6B6FE4 CRC64;

  Query Match             71.4%;  Score 1312;  DB 1;  Length 360;
  Best Local Similarity   69.2%;  Pred. No. 3.9e-86;
  Matches  249;  Conservative   44;  Mismatches   67;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: ||||| ||| || || |:| | || ::|||   || || ||| ||  |||| ||:|
Db          1 MLVYLAEYLTQFYSGFNVFSYVTFRAILGLMTALVFCLWWGPKMIRRLQTLQIGQVVRSD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| || ||||||| ||| | || ||| || :|||||| |   || ||::||||||
Db         61 GPESHFSKSGTPTMGGILILAGIFISVLLWGDLGSRYVWVVLFVLASFGLIGFIDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||| ||:  |  | :|| :|:|  || |:||  | :  |:| ||:||
Db        121 VRKDTKGLIARWKYILQSLAAIIIAFYLYASADTVGETQLVVPFFKDIMPQMGAFFIVLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              || |||||||||||||||||||||||| |  :  ||||:|:|| || || :||||||:
Db        181 YFTIVGSSNAVNLTDGLDGLAIMPTVMVAAAFALIAYLSGHVQFANYLHIPYLPGAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||:|||||||| |||:|||||:| |||||||||
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGAALGAIAVLVRQEILLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              |:||::|| |:|| |:|:|||||||||:||||||:||||||||:: ||::||||||||
Db        301 TVSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLFLVMLGLATLKLR 360
```

```
RESULT 7
MRAY_SHEON
ID   MRAY_SHEON     STANDARD;      PRT;   360 AA.
AC   Q8E9P5;
DT   15-MAR-2004 (Rel. 43, Created)
DT   15-MAR-2004 (Rel. 43, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR SO4222.
OS   Shewanella oneidensis.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC   Alteromonadaceae; Shewanella.
OX   NCBI_TaxID=70863;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MR-1;
RX   MEDLINE=22297686; PubMed=12368813;
RA   Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA   Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA   Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA   Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA   Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA   Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA   Feldblyum T.V., Smith H.O., Venter J.C., Nealson K.H., Fraser C.M.;
RT   "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT   Shewanella oneidensis.";
RL   Nat. Biotechnol. 20:1118-1123(2002).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
```

```
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE015855; AAN57194.1; -.
DR   TIGR; SO4222; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     20     42       Potential.
FT   TRANSMEM     74     93       Potential.
FT   TRANSMEM     97    114       Potential.
FT   TRANSMEM    134    153       Potential.
FT   TRANSMEM    168    190       Potential.
FT   TRANSMEM    197    219       Potential.
FT   TRANSMEM    234    256       Potential.
FT   TRANSMEM    263    285       Potential.
FT   TRANSMEM    289    311       Potential.
FT   TRANSMEM    338    357       Potential.
SQ   SEQUENCE   360 AA;  39783 MW;  CF447768C8F73217 CRC64;

  Query Match             71.1%;  Score 1306;  DB 1;  Length 360;
  Best Local Similarity   68.6%;  Pred. No. 1e-85;
  Matches  247;  Conservative   47;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: ||||| :|: || || |:| | || :|||| ||| || :|  ||: |||| ||||
Db          1 MLVYLAEYLTRFHTGFNVFSYVTFRAILGLLTALMFSLWWGPKLIERLQLMQIGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| ||| || :| ||| || :|||||:| |   || ||::|||||
Db         61 GPESHFSKRGTPTMGGLLILGAIFLSVLLWGDLGSRYVWVMLFVLGSFGMIGFIDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||| ||:  :  | ||| ||  | || |:||  | |  |||  |:||
Db        121 VRKDTKGLIARWKYILQSLAALIIAFFLYTTAANPGETQLVVPFFKDVMPQLGAVFIVLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              || |||||||||||||||||||||||| |  :  ||||: :|| || ||::||:|||:
Db        181 YFTIVGSSNAVNLTDGLDGLAIMPTVMVAAAFALIAYLSGHAQFANYLHIPHLPGSGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|:|||||| |||:|||||:| |||||||||
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLSLGAALGAIAVLVRQEILLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              |:||::|| |:|| |:|:|||||||||:||||||:|||||||||:: |||:||||||||
Db        301 TVSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISIFLVLLGLATLKLR 360


RESULT 8
MRAY_XYLFA
ID   MRAY_XYLFA     STANDARD;      PRT;   361 AA.
AC   Q9PF83;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR XF0795.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9a5c;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE003919; AAF83605.1; -.
DR   PIR; B82763; B82763.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     21     43       POTENTIAL.
FT   TRANSMEM     71     90       POTENTIAL.
```

```
FT   TRANSMEM     97    114       POTENTIAL.
FT   TRANSMEM    139    161       POTENTIAL.
FT   TRANSMEM    168    187       POTENTIAL.
FT   TRANSMEM    202    224       POTENTIAL.
FT   TRANSMEM    237    259       POTENTIAL.
FT   TRANSMEM    263    285       POTENTIAL.
FT   TRANSMEM    290    312       POTENTIAL.
FT   TRANSMEM    339    358       POTENTIAL.
SQ   SEQUENCE   361 AA;  39694 MW;  83A40BDD66F990CF CRC64;

  Query Match             70.6%;  Score 1296.5;  DB 1;  Length 361;
  Best Local Similarity   68.1%;  Pred. No. 4.9e-85;
  Matches  246;  Conservative   43;  Mismatches   71;  Indels    1;  Gaps    1;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||   || :||||    ||:| ||| | ||: |||| ||||:|| :|: |   : || :| |
Db          1 MLFELARWLQQFESLFGLFNYLTFRSILAALTALFLSLWIGPVLIQKLSQFKGGQPIRQD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||: | || |||||||:|||  : :| ||| || |||||:|||| | |||||| ||: |:
Db         61 GPKMHFSKAGTPTMGGSLILMTVTLSVLLWGDLRNRYVWLVLVVMLAFGAIGWYDDWIKL 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQL-GIFFVVL 179
               ::  || |||||  ||:||: | :||| ||: |   |  :|| ||: : |  | || :
Db        121 ARRDPNGLKSRWKYLLQSIFGLAAGLFLYFTADVPAAVTFYIPMFKSIALPLTSISFVAI 180

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
              ||| ||| |||||||||||||||||:|| |||:| | |||  |: || || :||||:|
Db        181 TYFWIVGFSNAVNLTDGLDGLAIMPTVLVACALGVFAYASGNTLFSSYLKIPTIPGAGDL 240

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
              |: |||: |||||||||| ||| |||||:|||||| ||||||||||:|| :||||||:
Db        241 IIICAAIAGAGLGFLWFNAYPAMVFMGDIGALALGAVLGTIAVIVRQELVLVVMGGVFVI 300

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              |||||:||| ||||||:||||||||||||||||:|||||||||:|:|||:||||||:
Db        301 ETLSVIIQVTSFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIISVVLVLVGLATLKV 360

Qy        360 R 360
              |
Db        361 R 361
```

```
RESULT 9
MRAY_VIBPA
ID   MRAY_VIBPA     STANDARD;      PRT;   360 AA.
AC   Q87SG7;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR VP0457.
OS   Vibrio parahaemolyticus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=670;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=RIMD 2210633 / Serotype O3:K6;
RX   MEDLINE=22508454; PubMed=12620739;
RA   Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA   Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA   Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT   "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT   distinct from that of V. cholerae.";
RL   Lancet 361:743-749(2003).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP005074; BAC58720.1; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     23     45       Potential.
FT   TRANSMEM     74     93       Potential.
FT   TRANSMEM     97    114       Potential.
FT   TRANSMEM    134    151       Potential.
FT   TRANSMEM    171    193       Potential.
FT   TRANSMEM    200    219       Potential.
FT   TRANSMEM    234    256       Potential.
FT   TRANSMEM    263    285       Potential.
FT   TRANSMEM    289    311       Potential.
FT   TRANSMEM    338    357       Potential.
SQ   SEQUENCE   360 AA;  39943 MW;  4E032FA3F11C6E01 CRC64;

  Query Match             70.4%;  Score 1293;  DB 1;  Length 360;
  Best Local Similarity   67.5%;  Pred. No. 8.6e-85;
  Matches  243;  Conservative   51;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |:: ||| || :   | :|:||: | |||||||| ||||:|| ||: ||: |||| |||:
Db          1 MIIWLAELLQPYLSFFRLFEYLSFRAILSVLTALGLSLWMGPIMIKRLQMLQIGQVVRNE 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| ||:|: ||| |||| ||| || | | :||:|:|||||||
Db         61 GPESHFSKRGTPTMGGIMILAAISITILLWTDLSNPYVWAVLTVLLGYGAVGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||: || :||||||||:      |  ||   :    | |:||  | |  |||: :::||
Db        121 VRKNTDGLIARWKYFWQSLIAFVVAFALYAYGKDTAATQLVVPFFKDVMPQLGLMYIILT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:||||||||||||||||||:||    :  : :||| |:||| || :| | ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAAGFAVIAWATGNVNFSEYLHIPYLPHASELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              | | |:|||||||||||||||||||||:|||| |||||||:||||:|| |||||||||
Db        241 VVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLVRQELVLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:|| |:|:||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISMVLVLIGLATLKVR 360
```

```
RESULT 10
MRAY_XYLFT
ID   MRAY_XYLFT     STANDARD;      PRT;   361 AA.
AC   Q87AF7;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR PD1868.
OS   Xylella fastidiosa (strain Temecula1 / ATCC 700964).
```

```
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC   Xanthomonadaceae; Xylella.
OX   NCBI_TaxID=183190;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22421331; PubMed=12533478;
RA   Van Sluys M.A., de Oliveira M.C., Monteiro-Vitorello C.B.,
RA   Miyaki C.Y., Furlan L.R., Camargo L.E.A., da Silva A.C.R., Moon D.H.,
RA   Takita M.A., Lemos E.G.M., Machado M.A., Ferro M.I.T., da Silva F.R.,
RA   Goldman M.H.S., Goldman G.H., Lemos M.V.F., El-Dorry H., Tsai S.M.,
RA   Carrer H., Carraro D.M., de Oliveira R.C., Nunes L.R., Siqueira W.J.,
RA   Coutinho L.L., Kimura E.T., Ferro E.S., Harakava R., Kuramae E.E.,
RA   Marino C.L., Giglioti E., Abreu I.L., Alves L.M.C., do Amaral A.M.,
RA   Baia G.S., Blanco S.R., Brito M.S., Cannavan F.S., Celestino A.V.,
RA   da Cunha A.F., Fenille R.C., Ferro J.A., Formighieri E.F., Kishi L.T.,
RA   Leoni S.G., Oliveira A.R., Rosa V.E. Jr., Sassaki F.T., Sena J.A.D.,
RA   de Souza A.A., Truffi D., Tsukumo F., Yanai G.M., Zaros L.G.,
RA   Civerolo E.L., Simpson A.J.G., Almeida N.F. Jr., Setubal J.C.,
RA   Kitajima J.P.;
RT   "Comparative analyses of the complete genome sequences of Pierce's
RT   disease and citrus variegated chlorosis strains of Xylella
RT   fastidiosa.";
RL   J. Bacteriol. 185:1018-1026(2003).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE012560; AAO29700.1; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     21     43       Potential.
FT   TRANSMEM     71     90       Potential.
FT   TRANSMEM     97    114       Potential.
FT   TRANSMEM    139    161       Potential.
FT   TRANSMEM    168    187       Potential.
FT   TRANSMEM    202    224       Potential.
FT   TRANSMEM    237    259       Potential.
FT   TRANSMEM    263    285       Potential.
FT   TRANSMEM    290    312       Potential.
FT   TRANSMEM    339    358       Potential.
SQ   SEQUENCE   361 AA;  39728 MW;  0EB87DACA0735E43 CRC64;

  Query Match             70.4%;  Score 1292.5;  DB 1;  Length 361;
  Best Local Similarity   67.9%;  Pred. No. 9.3e-85;
  Matches  245;  Conservative   43;  Mismatches   72;  Indels    1;  Gaps    1;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||    | :||||   ||:| ||| | ||: |||| ||||:|| :|: |   : || :| |
Db          1 MLFEFARWLQQFESLFGLFNYLTFRSILAALTALFLSLWIGPVLIQKLSQFKGGQPIRQD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||: | || |||||||:|||  : :| ||| || |||||:|||| | |||||| ||: |:
Db         61 GPKMHFSKAGTPTMGGSLILMTVTLSVLLWGDLRNRYVWLVLVVMLAFGAIGWYDDWIKL 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQL-GIFFVVL 179
              ::  || |||||  ||:||: | :||| ||: |   |  :|| ||: : |  | || :
Db        121 ARRDPNGLKSRWKYLLQSIFGLAAGLFLYFTADVPAAVTFYIPMFKSIALPLTSISFVAI 180

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
              ||| ||| |||||||||||||||||:|| |||:| | |||  |: || || :||||:|
Db        181 TYFWIVGFSNAVNLTDGLDGLAIMPTVLVACALGVFAYASGNTLFSSYLKIPTIPGAGDL 240

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
              |: |||: |||||||||| ||| |||||:|||||| ||||||||||:|| :||||||:
Db        241 IIICAAIAGAGLGFLWFNAYPAMVFMGDIGALALGAVLGTIAVIVRQELVLVVMGGVFVI 300

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              |||||:||| ||||||:||||||||||||||||:||||||||:|:|||:||||||:
Db        301 ETLSVIIQVTSFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIISVVLVLVGLATLKV 360

Qy        360 R 360
              |
Db        361 R 361


RESULT 11
MRAY_VIBCH
ID   MRAY_VIBCH     STANDARD;      PRT;   360 AA.
AC   Q9KPG4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR VC2404.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE004310; AAF95547.1; -.
DR   PIR; H82081; H82081.
DR   TIGR; VC2404; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
```

```
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM      22     41       POTENTIAL.
FT   TRANSMEM      74     93       POTENTIAL.
FT   TRANSMEM      97    114       POTENTIAL.
FT   TRANSMEM     134    151       POTENTIAL.
FT   TRANSMEM     171    193       POTENTIAL.
FT   TRANSMEM     200    222       POTENTIAL.
FT   TRANSMEM     237    256       POTENTIAL.
FT   TRANSMEM     263    285       POTENTIAL.
FT   TRANSMEM     289    311       POTENTIAL.
FT   TRANSMEM     338    357       POTENTIAL.
SQ   SEQUENCE   360 AA;  39819 MW;  81727653AF899E53 CRC64;

  Query Match             70.3%;  Score 1292;  DB 1;  Length 360;
  Best Local Similarity   66.9%;  Pred. No. 1e-84;
  Matches  241;  Conservative   51;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |:: ||| || ::  | :|:||: | |:|:|||| :|||:|| ||: ||: |||| |||:
Db          1 MIIWLAELLQPYFSFFRLFEYLSFRAIVSILTALGISLWMGPRMIKRLQMLQIGQVVRNE 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || |: |||||:| ||| || | | :||:|:|||||||
Db         61 GPESHFSKRGTPTMGGVMILAAITITVLLWADLTNPYVWAVLAVLLGYGAVGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||: || :|||||||    :  |  ||   :    | |:||  | |  |||: ::|||
Db        121 VRKNTDGLIARWKYFWQSAIALVVAFALYAHGQDTAATQLVVPFFKDVMPQLGLMYIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||:||       : :||| || || || :| : ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAAGFAAIAWATGNVNFANYLHIPYIPHSSELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              | | |:|||||||||||||||||||||||||||| |||||||:|||| || |||||||||
Db        241 VVCTAMVGAGLGFLWFNTYPAQVFMGDVGALALGGALGTIAVLVRQEFVLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:|| |:|:|||||||||:||||||:||||||||||:::||||||||||:|
Db        301 TLSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISIVLVLIGLATLKVR 360


RESULT 12
MRAY_VIBVU
ID   MRAY_VIBVU     STANDARD;      PRT;   360 AA.
AC   Q8DEK7;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR VV10581.
OS   Vibrio vulnificus.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC   Vibrionaceae; Vibrio.
OX   NCBI_TaxID=672;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CMCP6;
RA   Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA   Choy H.E.;
RT   "Complete genome sequence of Vibrio vulnificus CMCP6.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE016798; AAO09097.1; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM      15     37       POTENTIAL.
FT   TRANSMEM      74     93       POTENTIAL.
FT   TRANSMEM      97    114       POTENTIAL.
FT   TRANSMEM     134    151       POTENTIAL.
FT   TRANSMEM     171    193       POTENTIAL.
FT   TRANSMEM     200    219       POTENTIAL.
FT   TRANSMEM     234    256       POTENTIAL.
FT   TRANSMEM     263    285       POTENTIAL.
FT   TRANSMEM     289    311       POTENTIAL.
FT   TRANSMEM     338    357       POTENTIAL.
SQ   SEQUENCE   360 AA;  39850 MW;  D087052F04DBA0B1 CRC64;

  Query Match             70.3%;  Score 1291;  DB 1;  Length 360;
  Best Local Similarity   66.9%;  Pred. No. 1.2e-84;
  Matches  241;  Conservative   55;  Mismatches   64;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |:: ||| || ::  | :|:||: | |||||||| ||||:|| :|: ||: |||| |||:
Db          1 MIIWLAELLQPYFSFFRLFEYLSFRAILSVLTALGLSLWMGPRLIKRLQLLQIGQVVRNE 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || |: |||||:||| ||| || | : :||:|:|||||||
Db         61 GPESHFSKRGTPTMGGVMILAAISITILLWANLSNPYVWAVLAVLMGYGAVGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||: || :|||||||    :  |  ||   :    | |:||  | |  |||: ::|||
Db        121 VRKNTDGLIARWKYFWQSAIALIVAFALYAYGKDTAATQLVVPFFKDVMPQLGLMYIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||:||       : :||| |::|| || :| | ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAAGFAVIAWATGNVNFSQYLHIPYLPHASELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              | | |:|||||||||||||||||||||||||:|||| |||||||:||||:|| |||||||||
Db        241 VVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLVRQELVLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:|| |:|:|||||||||:||||||:||||||||||:::||||||||||:|
Db        301 TLSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISMVLVLIGLATLKVR 360


RESULT 13
MRAY_PASMU
ID   MRAY_PASMU     STANDARD;      PRT;   360 AA.
AC   P57816;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
```

```
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR PM0139.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC   Pasteurellaceae; Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Pm70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE006048; AAK02223.1; -.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM     21     41       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM     94    114       POTENTIAL.
FT   TRANSMEM    132    152       POTENTIAL.
FT   TRANSMEM    168    188       POTENTIAL.
FT   TRANSMEM    200    220       POTENTIAL.
FT   TRANSMEM    236    256       POTENTIAL.
FT   TRANSMEM    263    283       POTENTIAL.
FT   TRANSMEM    288    308       POTENTIAL.
FT   TRANSMEM    338    358       POTENTIAL.
SQ   SEQUENCE   360 AA;  40155 MW;  B9BED1987FEDBD53 CRC64;

  Query Match             69.4%;  Score 1274;  DB 1;  Length 360;
  Best Local Similarity   65.8%;  Pred. No. 1.9e-83;
  Matches  237;  Conservative   57;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||:  |:|||:| ||  |  |:|:| ||::|||| :|||:|| :|| ||: : || ||:|
Db          1 MLVWLGEFLQQYYSGFNVISYITVRAILALLTALLVSLWIGPKVIRRLQLLKFGQEVRHD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || :||||||:|:| |||  | | | :||||:||||||:
Db         61 GPESHFSKRGTPTMGGVMILFAITVSTLLWANLANPYVWFSLFVLLGYGAIGFVDDYRKI 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
                ||: || :|||||| ||  : ||  :|   :    | |:||  | :  |||:|:::||
Db        121 TRKNTDGLIARWKYFWLSVIALVAAFGMYAIGKDTDATRLVVPFFKEIMPQLGLFYIILT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||:|||| :  : :|||  ||||| || :  : ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAGAFALIAWATGNVNFAEYLHIPYIKFSAELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              |||  |:|||||||||||||||||||:|||| |||  :||:||||  :|  |||||||:|
Db        241 VFCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGVVAVLVRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:||  :|:|||||||||||||||:||||||||:::|||:|| |||||
Db        301 TLSVILQVGSYKLRKQRIFRMAPIHHHFELKGWPEPRVIVRFWIISLMLVLVGLVTLKLR 360


RESULT 14
MRAY_ECOL6
ID   MRAY_ECOL6     STANDARD;      PRT;   360 AA.
AC   Q8X9Z0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR C0105 OR Z0097 OR ECS0091.
OS   Escherichia coli O6, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=217992, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX   MEDLINE=22388234; PubMed=12471157;
RA   Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA   Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA   Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA   Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT   "Extensive mosaic structure revealed by the complete genome sequence
RT   of uropathogenic Escherichia coli.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan (By similarity).
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; AE016755; AAN78603.1; -.
DR   EMBL; AE005185; AAG54391.1; -.
DR   EMBL; AP002550; BAB33514.1; -.
DR   PIR; C85491; C85491.
DR   PIR; C90640; C90640.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Complete proteome.
FT   DOMAIN       1     18       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM     19    45       BY SIMILARITY.
FT   DOMAIN       46    76       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     77    90       BY SIMILARITY.
FT   DOMAIN       91    96       PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM     97    113      BY SIMILARITY.
FT   DOMAIN       114   133      CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     134   156      BY SIMILARITY.
FT   DOMAIN       157   173      PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM     174   188      BY SIMILARITY.
FT   DOMAIN       189   199      CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     200   220      BY SIMILARITY.
FT   DOMAIN       221   238      PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM     239   251      BY SIMILARITY.
FT   DOMAIN       252   270      CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     271   284      BY SIMILARITY.
FT   DOMAIN       285   287      PERIPLASMIC (BY SIMILARITY).
FT   TRANSMEM     288   299      BY SIMILARITY.
FT   DOMAIN       300   341      CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     342   357      BY SIMILARITY.
FT   DOMAIN       358   360      PERIPLASMIC (BY SIMILARITY).
SQ   SEQUENCE   360 AA;  39903 MW;  922AB374BC9E8F2E2 CRC64;

  Query Match             69.0%;  Score 1268;  DB 1;  Length 360;
  Best Local Similarity   66.4%;  Pred. No. 5.1e-83;
  Matches  239;  Conservative   51;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:|  ::| || || ||| | |:|:|||| :|||:|| ||   ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :||||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||: |||||||||||||||||| |||   :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 15
MRAY_ECOLI
ID   MRAY_ECOLI     STANDARD;      PRT;   360 AA.
AC   P15876;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) (UDP-
DE   MurNAc-pentapeptide phosphotransferase).
GN   MRAY OR MURX OR B0087 OR SF0084 OR S0086.
OS   Escherichia coli, and
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
OX   NCBI_TaxID=562, 623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12;
RX   MEDLINE=90192099; PubMed=2179861;
RA   Ikeda M., Wachi M., Ishino F., Matsuhashi M.;
RT   "Nucleotide sequence involving murD and an open reading frame ORF-Y
RT   spacing murF and ftsW in Escherichia coli.";
RL   Nucleic Acids Res. 18:1058-1058(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12;
RX   MEDLINE=92334977; PubMed=1630901;
RA   Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA   Isono K., Mizobuchi K., Nakata A.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of
RT   the 0-2.4 min region.";
RL   Nucleic Acids Res. 20:3305-3308(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;
RX   MEDLINE=22272406; PubMed=12384590;
RA   Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA   Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA   Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA   Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA   Yu J.;
RT   "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT   through comparison with genomes of Escherichia coli K12 and O157.";
RL   Nucleic Acids Res. 30:4432-4441(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX   MEDLINE=22590274; PubMed=12704152;
RA   Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA   Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA   Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA   Schwartz D.C., Blattner F.R.;
RT   "Complete genome sequence and comparative genomics of Shigella
RT   flexneri serotype 2a strain 2457T.";
RL   Infect. Immun. 71:2775-2786(2003).
RN   [6]
RP   CHARACTERIZATION.
RC   SPECIES=E.coli; STRAIN=K12;
RX   MEDLINE=79082893; PubMed=215212;
RA   Geis A., Plapp R.;
RT   "Phospho-N-acetylmuramoyl-pentapeptide-transferase of Escherichia coli
RT   K12. Properties of the membrane-bound and the extracted and partially
RT   purified enzyme.";
RL   Biochim. Biophys. Acta 527:414-424(1978).
RN   [7]
RP   FUNCTION.
RC   SPECIES=E.coli; STRAIN=K12;
```

```
RX   MEDLINE=91123172; PubMed=1846850;
RA   Ikeda M., Wachi M., Jung H.K., Ishino F., Matsuhashi M.;
RT   "The Escherichia coli mraY gene encoding UDP-N-acetylmuramoyl-
RT   pentapeptide: undecaprenyl-phosphate phospho-N-acetylmuramoyl-
RT   pentapeptide transferase.";
RL   J. Bacteriol. 173:1021-1026(1991).
RN   [8]
RP   TOPOLOGY.
RC   SPECIES=E.coli;
RX   MEDLINE=20032370; PubMed=10564498;
RA   Bouhss A., Mengin-Lecreulx D., Le Beller D., Van Heijenoort J.;
RT   "Topological analysis of the mraY protein catalyzing the first
RT   membrane step of peptidoglycan synthesis.";
RL   Mol. Microbiol. 34:576-585(1999).
CC   -!- FUNCTION: First step of the lipid cycle reactions in the
CC       biosynthesis of the cell wall peptidoglycan.
CC   -!- CATALYTIC ACTIVITY: UDPMur2Ac(oyl-L-Ala-gamma-D-Glu-L-Lys-D-Ala-D-
CC       Ala) + undecaprenyl phosphate = UMP + Mur2Ac(oyl-L-Ala-gamma-D-
CC       Glu-L-Lys-D-Ala-D-Ala)-diphosphoundecaprenol.
CC   -!- COFACTOR: Magnesium or manganese.
CC   -!- PATHWAY: Peptidoglycan biosynthesis.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: Belongs to the glycosyltransferase family 4. MraY
CC       subfamily.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51584; CAA35932.1; -.
DR   EMBL; X55034; CAA38864.1; -.
DR   EMBL; D10483; BAB96655.1; -.
DR   EMBL; AE000118; AAC73198.1; -.
DR   EMBL; AE015046; AAN41749.1; -.
DR   EMBL; AE016978; AAP15630.1; -.
DR   PIR; S08395; S08395.
DR   EcoGene; EG10604; mraY.
DR   HAMAP; MF_00038; -; 1.
DR   InterPro; IPR000715; Glyco_trans_4.
DR   InterPro; IPR003524; PNAcPpept_trans.
DR   Pfam; PF00953; Glycos_transf_4; 1.
DR   TIGRFAMs; TIGR00445; mraY; 1.
DR   PROSITE; PS01347; MRAY_1; 1.
DR   PROSITE; PS01348; MRAY_2; 1.
KW   Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
KW   Magnesium; Manganese; Complete proteome.
FT   DOMAIN        1     18       PERIPLASMIC.
FT   TRANSMEM     19     45
FT   DOMAIN       46     76       CYTOPLASMIC.
FT   TRANSMEM     77     90
FT   DOMAIN       91     96       PERIPLASMIC.
FT   TRANSMEM     97    113
FT   DOMAIN      114    133       CYTOPLASMIC.
FT   TRANSMEM    134    156
FT   DOMAIN      157    173       PERIPLASMIC.
FT   TRANSMEM    174    188
FT   DOMAIN      189    199       CYTOPLASMIC.
FT   TRANSMEM    200    220
FT   DOMAIN      221    238       PERIPLASMIC.
FT   TRANSMEM    239    251
FT   DOMAIN      252    270       CYTOPLASMIC.
FT   TRANSMEM    271    284
FT   DOMAIN      285    287       PERIPLASMIC.
FT   TRANSMEM    288    299
FT   DOMAIN      300    341       CYTOPLASMIC.
FT   TRANSMEM    342    357
FT   DOMAIN      358    360       PERIPLASMIC.
SQ   SEQUENCE   360 AA;  39874 MW;  F3550AFA3CD636AE CRC64;
```

```
  Query Match             69.0%;  Score 1268;  DB 1;  Length 360;
  Best Local Similarity   66.4%;  Pred. No. 5.1e-83;
  Matches  239;  Conservative   51;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :||||| || |||| || ||| |||| : :| ||:||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


Search completed: May  7, 2004, 08:38:11
Job time : 19 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 08:37:38 ; Search time 21 Seconds
(without alignments)
1648.998 Million cell updates/sec

Title: US-10-089-787-2
Perfect score: 1837
Sequence: 1 MLLLLAEYLQQFYKGFGVFQ.....RFWIITVILVLIGLATLKLR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1824	99.3	360	2	H83094	phospho-N-acetylmu
2	1296.5	70.6	361	2	B82763	phospho-N-acetylmu
3	1292	70.3	360	2	H82081	phospho-N-acetylmu
4	1268	69.0	360	1	S08395	phospho-N-acetylmu
5	1268	69.0	360	2	C90640	hypothetical prote
6	1268	69.0	360	2	C85491	hypothetical prote
7	1265	68.9	360	2	AE0068	phospho-N-acetylmu
8	1264	68.8	360	2	AB0518	phospho-N-acetylmu
9	1216	66.2	360	1	A64185	phospho-N-acetylmu
10	1100	59.9	360	2	E81200	phospho-N-acetylmu
11	1099	59.8	360	2	F81777	phospho-N-acetylmu
12	945	51.4	360	2	AB3324	phospho-N-acetylmu
13	917	49.9	366	2	AB2834	hypothetical prote
14	917	49.9	366	2	F97611	phospho-N-acetylmu
15	915.5	49.8	357	2	G84955	phospho-N-acetylmu
16	890.5	48.5	361	2	F97813	hypothetical prote
17	887.5	48.3	361	2	E71664	phospho-n-acetylmu
18	853.5	46.5	359	2	F70304	phospho-N-acetylmu
19	812.5	44.2	351	2	G70137	phospho-N-acetylmu
20	782.5	42.6	353	2	G81387	phospho-N-acetylmu
21	709.5	38.6	353	2	B71930	phospho-n-acetylmu
22	708.5	38.6	353	1	E64581	phospho-N-acetylmu
23	652	35.5	321	2	A89890	phospho-N-muramic
24	651	35.4	324	1	C47691	phospho-N-acetylmu
25	639	34.8	324	2	AE1329	phospho-N-acetylmu
26	636	34.6	324	2	AE1700	phospho-N-acetylmu
27	636	34.6	363	2	T34957	probable phospho-N
28	605.5	33.0	317	2	B97162	phospho-N-acetylmu
29	594.5	32.4	369	2	AE2345	phospho-N-acetylmu
30	555.5	30.2	247	2	H83970	phospho-N-acetylmu
31	551	30.0	359	2	H70579	probable murX prot
32	548	29.8	359	2	A87023	hypothetical prote
33	546.5	29.7	349	2	B86603	muramoyl-pentapept
34	546.5	29.7	349	2	A81519	phospho-N-acetylmu
35	543.5	29.6	349	2	G72021	muramoyl-pentapept
36	526.5	28.7	326	2	G95039	hypothetical prote
37	520	28.3	326	2	A97910	phospho-N-acetylmu
38	519	28.3	365	1	S77076	phospho-N-acetylmu
39	507	27.6	329	2	D86733	hypothetical prote
40	505.5	27.5	335	2	E81736	phospho-N-acetylmu
41	503.5	27.4	302	2	E72402	phospho-N-acetylmu
42	477.5	26.0	336	2	D71474	probable muramoyl-
43	449.5	24.5	312	2	F75346	phospho-N-acetylmu
44	342	18.6	236	2	T04929	hypothetical prote
45	295	16.1	351	1	C69198	phospho-N-acetylmu

ALIGNMENTS

RESULT 1
H83094
phospho-N-acetylmuramoyl-pentapeptide- transferase PA4415 [imported] - Pseudomonas aerug:
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83094
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br:
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83094
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <STO>
A;Cross-references: GB:AE004856; GB:AE004091; NID:g9950633; PIDN:AAG07803.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: mraY; PA4415
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

Query Match 99.3%; Score 1824; DB 2; Length 360;
Best Local Similarity 99.4%; Pred. No. 1.9e-130;
Matches 358; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              |||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 2

B82763
phospho-N-acetylmuramoyl-pentapeptide- transferase XF0795 [imported] - Xylella fastidios
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: B82763
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-361 <SIM>
A;Cross-references: GB:AE003919; GB:AE003849; NID:g9105675; PIDN:AAF83605.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laign
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF0795
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

Query Match 70.6%; Score 1296.5; DB 2; Length 361;
Best Local Similarity 68.1%; Pred. No. 1.2e-90;
Matches 246; Conservative 43; Mismatches 71; Indels 1; Gaps 1;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||   || :||||    ||:| ||| | ||: |||| ||||:|| :|: |    : || :| |
Db          1 MLFELARWLQQFESLFGLFNYLTFRSILAALTALFLSLWIGPVLIQKLSQFKGGQPIRQD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||: | || |||||||:|||  : :| ||| || |||||:|||| | |||||| ||: |:
Db         61 GPKMHFSKAGTPTMGGSLILMTVTLSVLLWGDLRNRYVWLVLVVMLAFGAIGWYDDWIKL 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQL-GIFFVVL 179
                 ::  || |||||  ||:||: | :||| ||:  |    |  :|| ||: : |  | || :
Db        121 ARRDPNGLKSRWKYLLQSIFGLAAGLFLYFTADVPAAVTFYIPMFKSIALPLTSISFVAI 180

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
              ||| ||| |||||||||||||||||||:|| |||:| | |||  |: || || :||||:|
Db        181 TYFWIVGFSNAVNLTDGLDGLAIMPTVLVACALGVFAYASGNTLFSSYLKIPTIPGAGDL 240

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
              |: |||: |||||||||| ||| |||||:|||||| ||||||||||||:|| :||||||:
Db        241 IIICAAIAGAGLGFLWFNAYPAMVFMGDIGALALGAVLGTIAVIVRQELVLVVMGGVFVI 300

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              |||||:||| ||||||:||||||||||||||||:|||||||||:|:|:||:||||||:
Db        301 ETLSVIIQVTSFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIISVVLVLVGLATLKV 360

Qy        360 R 360
              |
Db        361 R 361
```

RESULT 3
H82081
phospho-N-acetylmuramoyl-pentapeptide- transferase VC2404 [imported] - Vibrio cholerae
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: H82081
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: H82081
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <HEI>
A;Cross-references: GB:AE004310; GB:AE003852; NID:g9656963; PIDN:AAF95547.1; GSPDB:GN001
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC2404
A;Map position: 1
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

Query Match 70.3%; Score 1292; DB 2; Length 360;
Best Local Similarity 66.9%; Pred. No. 2.7e-90;
Matches 241; Conservative 51; Mismatches 68; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |:: ||| || ::  | :|:||: | |:|:|||| :|||:|| ||: ||: |||| |||:
Db          1 MIIWLAELLQPYFSFFRLFEYLSFRAIVSILTALGISLWMGPRMIKRLQMLQIGQVVRNE 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || |: ||||||:| ||| || | | :||:|:|||||||
Db         61 GPESHFSKRGTPTMGGVMILAAITITVLLWADLTNPYVWAVLAVLLGYGAVGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||: || :|||||||    :  |  ||   :    | |:||  | |  |||: ::|||
Db        121 VRKNTDGLIARWKYFWQSAIALVVAFALYAHGQDTAATQLVVPFFKDVMPQLGLMYIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||:||     : :||| || || || :| : ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAAGFAAIAWATGNVNFANYLHIPYIPHSSELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              | | |:|||||||||||||||||||||||||| |||||||:|||| || |||||||||
Db        241 VVCTAMVGAGLGFLWFNTYPAQVFMGDVGALALGGALGTIAVLVRQEPVLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:|| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISIVLVLIGLATLKVR 360
```

RESULT 4
S08395
phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) mraY - Escherichia coli
C;Species: Escherichia coli
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Mar-2002
C;Accession: S08395; S40597; G64730
R;Ikeda, M.; Wachi, M.; Ishino, F.; Matsuhashi, M.
Nucleic Acids Res. 18, 1058, 1990
A;Title: Nucleotide sequence involving murD and an open reading frame ORF-Y spacing murF
A;Reference number: S08395; MUID:90192099; PMID:2179861
A;Accession: S08395
A;Molecule type: DNA
A;Residues: 1-360 <IKE>
A;Cross-references: EMBL:X51584; NID:g42058; PIDN:CAA35932.1; PID:g42059
R;Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu
submitted to the EMBL Data Library, December 1992
A;Description: Systematic sequencing of the Escherichia coli genome: analysis of the 0-2
A;Reference number: S40531
A;Accession: S40597
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <YUR>
A;Cross-references: EMBL:D10483; NID:g216434; PIDN:BAA01352.1; PID:g216501
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: G64730
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-360 <BLAT>
A;Cross-references: GB:AE000118; GB:U00096; NID:g1786262; PIDN:AAC73198.1; PID:g1786275;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: mraY
A;Map position: 2 min
C;Function:
A;Pathway: peptidoglycan biosynthesis
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase
C;Keywords: cell division; peptidoglycan biosynthesis; transferase; transmembrane protei
F;22-38/Domain: transmembrane #status predicted <TM1>
F;73-89/Domain: transmembrane #status predicted <TM2>
F;98-114/Domain: transmembrane #status predicted <TM3>
F;135-151/Domain: transmembrane #status predicted <TM4>
F;172-188/Domain: transmembrane #status predicted <TM5>
F;200-216/Domain: transmembrane #status predicted <TM6>
F;239-255/Domain: transmembrane #status predicted <TM7>
F;268-284/Domain: transmembrane #status predicted <TM8>
F;289-305/Domain: transmembrane #status predicted <TM9>
F;338-354/Domain: transmembrane #status predicted <TM10>

Query Match 69.0%; Score 1268; DB 1; Length 360;
Best Local Similarity 66.4%; Pred. No. 1.7e-88;
Matches 239; Conservative 51; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || ||  :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||:|||| ||| ||| ||:|||  :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 5
C90640
hypothetical protein ECs0091 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: C90640
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90640
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB33514.1; PID:g13359547; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs0091
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

Query Match 69.0%; Score 1268; DB 2; Length 360;
Best Local Similarity 66.4%; Pred. No. 1.7e-88;
Matches 239; Conservative 51; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || ||  :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||:|||| ||| ||| ||:|||  :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 6
C85491
hypothetical protein mraY [imported] - Escherichia coli (strain O157:H7, substrain EDL93
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: C85491
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: C85491
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <STO>
A;Cross-references: GB:AE005174; NID:g12512790; PIDN:AAG54391.1; GSPDB:GN00145; UWGP:Z00
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: mraY
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

Query Match 69.0%; Score 1268; DB 2; Length 360;
Best Local Similarity 66.4%; Pred. No. 1.7e-88;
Matches 239; Conservative 51; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || ||  :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
```

```
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360

Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 7
AE0068
phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) [imported] - Yersinia pe
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C;Accession: AE0068
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AE0068
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC89408.1; PID:g15978644; GSPDB:GN00175
C;Genetics:
A;Gene: mraY
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase
C;Keywords: transferase

  Query Match             68.9%;  Score 1265;  DB 2;  Length 360;
  Best Local Similarity   65.6%;  Pred. No. 2.9e-88;
  Matches  236;  Conservative   54;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60

Db          1 MLVWLAEYLVKFYSGFNVFSYLTFRAIVSLLTALFISLWMGPHLIAWLQKLQIGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Db         61 GPESHFSKRGTPTMGGLMILFSITISVLMWAYPSNPYVWCVLFILIGYGIVGFIDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Db        121 VRKNTKGLIARWKYFWQSIIALAAAFTMYSIGKDTSATELVVPFFKDIMPQLGLLYVLLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNVNFAAYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360

Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 8
AB0518
phospho-N-acetylmuramoyl-pentapeptide- transferase [imported] - Salmonella enterica subs
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0518
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0518
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD01282.1; PID:g16501410; GSPDB:GN00176
C;Genetics:
A;Gene: STY0145
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

  Query Match             68.8%;  Score 1264;  DB 2;  Length 360;
  Best Local Similarity   66.1%;  Pred. No. 3.5e-88;
  Matches  238;  Conservative   52;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60

Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360

Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 9
A64185
phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) - Haemophilus influenzae
C;Species: Haemophilus influenzae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A64185
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A;Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: A64185
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-360 <TIGR>
A;Cross-references: GB:U32793; GB:L42023; NID:g1574683; PIDN:AAC22790.1; PID:g1574690; T
C;Genetics:
A;Gene: mraY
C;Function:
A;Pathway: peptidoglycan biosynthesis
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase
C;Keywords: cell division; peptidoglycan biosynthesis; transferase; transmembrane protei
F;29-45/Domain: transmembrane #status predicted <TM1>
F;77-93/Domain: transmembrane #status predicted <TM2>
F;98-114/Domain: transmembrane #status predicted <TM3>
F;135-151/Domain: transmembrane #status predicted <TM4>
F;172-188/Domain: transmembrane #status predicted <TM5>
F;200-216/Domain: transmembrane #status predicted <TM6>
F;239-255/Domain: transmembrane #status predicted <TM7>
F;268-284/Domain: transmembrane #status predicted <TM8>
F;289-305/Domain: transmembrane #status predicted <TM9>
F;338-354/Domain: transmembrane #status predicted <TM10>
```

```
  Query Match             66.2%;  Score 1216;  DB 1;  Length 360;
  Best Local Similarity   61.9%;  Pred. No. 1.5e-84;
  Matches  223;  Conservative   63;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: ||||| ::    |    |:|:|  |::|||| :|||:|| :|: ||| : || ||||
Db          1 MLVWLAEYLVRYETAFNAISYITVRANLALLTALFISLWIGPKVIKRLQILKFGQEVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| :||||||||| :|| :| :|||||:|:| |:|| | |   :||||:|||:||:
Db         61 GPESHFAKKGTPTMGGVMILFSIGVSTLLWANLANPYIWVCLFVLFGYGAIGFVDDFRKI 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
                ||: || :|||||| ||   : | ::||        | |::|  | :  |||:|::||:
Db        121 TRKNTDGLIARWKYFWMSVVALVAILWLYWLGHDTDATRLVIPFFKDIMPQLGLFYIVLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| :|||| :  : :||| ||||| || :  : |::
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTALVAGAFALIAWATGNVNFAEYLHIPYIKYSSEVV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||| |:||| |||||||||||||||||:|||| ||| :|::|||| :| |||||||:|
Db        241 VFCTAIVGASLGFLWFNTYPAQVFMGDVGSLALGGALGVVAILVRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              |||::|| |:||  :|:||||||||||||||:||||:|||||:::|||:|| |||||
Db        301 ALSVILQVGSYKLRKQRIFRMAPIHHHFELKGWPEPRVIIRFWIISLMLVLMGLVTLKLR 360


RESULT 10
E81200
phospho-N-acetylmuramoyl-pentapeptide- transferase NMB0418 [imported] - Neisseria mening
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: E81200
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: E81200
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <TET>
A;Cross-references: GB:AE002398; GB:AE002098; NID:g7225640; PIDN:AAF40856.1; PID:g722564
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB0418
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

  Query Match             59.9%;  Score 1100;  DB 2;  Length 360;
  Best Local Similarity   58.2%;  Pred. No. 8.4e-76;
  Matches  210;  Conservative   59;  Mismatches   90;  Indels    2;  Gaps    2;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              | | || :   :  |  :||| | | ::: |||| || ||| || |   : ||||| |
Db          1 MFLWLAHF-SNWLTGLNIFQYTTFRAVMAALTALAFSLMFGPWTIRRLTALKCGQAVRTD 59

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||:|| | ||||||:|||||| :||||| : :| |:|::| | |  ||:|: ||:|||
Db         60 GPQTHLVKNGTPTMGGSLILTAITVSTLLWGNWANPYIWILLGVLLATGALGFYDDWRKV 119

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIF-FVVL 179
              : |:  |: :::|  |||   | |:: |:  |       ||||  | : : ||:  |:||
Db        120 VYKDPNGVSAKFKMVWQSSVAIIASLALFYLAANSANNILIVPFFKQIALPLGVVGFLVL 179

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
              :|  |||:|||||||||||  | |:||  | || | ||: :||:||  :| | || |:
Db        180 SYLTIVGTSNAVNLTDGLDGLATFPVVLVAAGLAIFAYASGHSQFAQYLQLPYVAGANEV 239

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
```

```
Db        240 VIFCTAMCGACLGFLWFNAYPAQVFMGDVGALALGAALGTVAVIVRQEFVLVIMGGLFVV 299

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              | :|||:||  :| | :|:| |||||||:| ||| : :|:|||||||::|||||:|||:
Db        300 EAVSVMLQVGWYKKTKKRIFLMAPIHHHYEQKGWKETQVVVRFWIITIVLVLIGLSTLKI 359

Qy        360 R 360
              |
Db        360 R 360


RESULT 11
F81777
phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) NMA2066 [imported] - Neis
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: F81777
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: F81777
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <PAR>
A;Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85284.1; PID:g738069
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: mraY; NMA2066
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase
C;Keywords: transferase

  Query Match             59.8%;  Score 1099;  DB 2;  Length 360;
  Best Local Similarity   58.2%;  Pred. No. 1e-75;
  Matches  210;  Conservative   58;  Mismatches   91;  Indels    2;  Gaps    2;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              | | ||  :   :  |  :||| | | ::: |||| || ||| || |   : ||||| |
Db          1 MFLWLAHF-SNWLTGLNIFQYTTFRAVMAALTALAFSLMFGPWTIRRLTALKCGQAVRTD 59

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||:|| | |||||||:|||||| :||||| : :| |:|::| | |  ||:|: ||:|||
Db         60 GPQTHLVKNGTPTMGGSLILTAITVSTLLWGNWANPYIWILLGVLLATGALGFYDDWRKV 119

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIF-FVVL 179
              : |:  |: :::|  |||   | |:: |:  |       ||||  | : : ||:  |:||
Db        120 VYKDPNGVSAKFKMVWQSSVAIIAGLALFYLAANSANNILIVPFFKQIALPLGVVGFLVL 179

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
              :|  |||:|||||||||||  | |:||  | || | ||: :||:||  :| | || |:
Db        180 SYLTIVGTSNAVNLTDGLDGLATFPVVLVAAGLAIFAYASGHSQFAQYLQLPYVAGANEV 239

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
              ::|| |: || ||||||| |||||||||||||||||||:||||||| || ||||:||:
Db        240 VIFCTAMCGACLGFLWFNAYPAQVFMGDVGALALGAALGTVAVIVRQEFVLVIMGGLFVV 299

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              | :|||:||  :| | :|:| |||||||:| ||| : :|:|||||||::|||||:|||:
Db        300 EAVSVMLQVGWYKKTKKRIFLMAPIHHHYEQKGWKETQVVVRFWIITIVLVLIGLSTLKI 359

Qy        360 R 360
              |
Db        360 R 360


RESULT 12
AB3324
phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) [imported] - Brucella mel
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 19-Apr-2002
C;Accession: AB3324
```

```
R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AB3324
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-360 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL51757.1; PID:g17982496; GSPDB:GN00190
A;Experimental source: strain 16M
C;Genetics:
A;Gene: BMEI0576
A;Map position: I
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase
C;Keywords: transferase

  Query Match             51.4%;  Score 945;  DB 2;  Length 360;
  Best Local Similarity   51.2%;  Pred. No. 4.2e-64;
  Matches  185;  Conservative   65;  Mismatches  109;  Indels    2;  Gaps    2;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQ-IGQAVRN 59
              ||: |  : :     | ||:|:| |   :::|:  :    || :| :|:: |  || :|
Db          1 MLMFLTHFAEH-VTPFNVFRYITFRTGGAMITSALIVFLFGPTIINSLRVRQGKGQPIRA 59

Qy         60 DGPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRK 119
              ||||:|  | |||||| :|:| |  | ||||:|:: ||||||:|:: |||||: ||| |
Db         60 DGPQTHFKKAGTPTMGGLMIMTGILASCLLWANLASVYVWVVLMVSVGFGAIGFYDDYLK 119

Qy        120 VIEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVL 179
              | ::: :|   : :   : :    || :   : |  ::|  | :| : | |  ||:
Db        120 VTKQSDKGFSGKARLGIEFLIAAIAAFTIMRAGQEPFSSSLTFPFVKQLVINLSWFFIPF 179

Qy        180 TYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGEL 239
              ||:||: |||||||||||||||:| :: | : |   |||||  ||:|| |  ||| |||
Db        180 AAFVMVGAGNAVNLTDGLDGLAIVPVMVAAASFGFIAYLSGNAIFADYLQIHFVPGTGEL 239

Qy        240 IVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVM 299
              |   |::|||||||||||  || :|||| |:|||  |||:||  : |||| |:||:|||
Db        240 AVVLGAVIGAGLGFLWFNAPPAAIFMGDTGSLALGGMLGTVAVATKHEIVLAIIGGLFVM 299

Qy        300 ETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359
              | |||:|||  ||:|||||| |||||||||| ||| : :|::||||: :|| :|||:||||
Db        300 EALSVIIQVGFFKMTGRRVFLMAPIHHHFEKKGWTESQVVIRFWIVAIILAMIGLSTLKL 359

Qy        360 R 360
              |
Db        360 R 360


RESULT 13
AB2834
hypothetical protein mraY [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C;Accession: AB2834
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; MUID:21608550; PMID:11743193
A;Accession: AB2834
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43088.1; PID:g17740558; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: mraY
A;Map position: circular chromosome
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

  Query Match             49.9%;  Score 917;  DB 2;  Length 366;
  Best Local Similarity   52.0%;  Pred. No. 5.6e-62;
  Matches  194;  Conservative   53;  Mismatches  106;  Indels   20;  Gaps    6;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQ-IGQAVRN 59
              ||: | | |    : | :|:|:| |   :: |:  :    || :| :|:: |  || :|
Db          1 MLIWLVE-LSDKVQLFNLFRYITFRAGAAMFTSALIVFLFGPAIINSLRVRQGKGQPIRA 59

Qy         60 DGPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRK 119
              ||||:|  | |||||| :|| |   :||| |||| ||  ||:||| |||||: ||| |
Db         60 DGPQTHFKKAGTPTMGGLMILAGILGGSLLWGDLSNVYVVAVLMVTLGFGAIGFYDDYLK 119

Qy        120 VIEKNSRGLPSRWKYFWQSVFGI-----GAAVFLYM---TAETP----IETTLIVPMLKS 167
              | ::: :|   : :       ||       ||| |     | |     : :::  | |
Db        120 VTKQSDKGFSGKAR------LGIEFLIAAIAVFFMMKMALASAPHGGTLGSSIAFPFFKE 173

Qy        168 VEIQLGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEY 227
              | || |||:   |||||: ||||||||||||||:| :: |   |:  ||:||  || |
Db        174 FVINLGYFFVLFGAFVIVGAGNAVNLTDGLDGLAIVPVMIAAATFGVIAYLAGNAVFANY 233

Qy        228 LLIPNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQE 287
              | |  ||| ||| |   |::|||||||||  || :|||| |:|||| :|:|||  : |
Db        234 LQINFVPGTGELAVIVGAVIGAGLGFLWFNAPPAAIFMGDTGSLALGGLIGSIAVATKHE 293

Qy        288 IVLFIMGGVFVMETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITV 347
              ||: |:||:|||||||:||| || |||||| |||||||| ||| : :|::|||||:|
Db        294 IVMVIVGGLFVMETLSVIIQVFWFKRTGRRVFLMAPIHHHFEKKGWTESQVVIRFWIISV 353

Qy        348 ILVLIGLATLKLR 360
              | |:||||||||
Db        354 GLALLGLATLKLR 366


RESULT 14
F97611
phospho-N-acetylmuramoyl-pentapeptide-transferase (UDP-murnac-pentapeptide phosphotransf
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C;Accession: F97611
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: F97611
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-366 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87847.1; PID:g15157231; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_3805
A;Map position: circular chromosome
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase

  Query Match             49.9%;  Score 917;  DB 2;  Length 366;
  Best Local Similarity   52.0%;  Pred. No. 5.6e-62;
  Matches  194;  Conservative   53;  Mismatches  106;  Indels   20;  Gaps    6;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQ-IGQAVRN 59
              ||: | | |    : | :|:|:| |   :: |:  :    || :| :|:: |  || :|
Db          1 MLIWLVE-LSDKVQLFNLFRYITFRAGAAMFTSALIVFLFGPAIINSLRVRQGKGQPIRA 59

Qy         60 DGPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRK 119
              ||||:|  | |||||| :|| |   :||| |||| ||  ||:||| |||||: ||| |
Db         60 DGPQTHFKKAGTPTMGGLMILAGILGGSLLWGDLSNVYVVAVLMVTLGFGAIGFYDDYLK 119

Qy        120 VIEKNSRGLPSRWKYFWQSVFGI-----GAAVFLYM---TAETP----IETTLIVPMLKS 167
              | ::: :|   : :       ||       ||| |     | |     : :::  | |
Db        120 VTKQSDKGFSGKAR------LGIEFLIAAIAVFFMMKMALASAPHGGTLGSSIAFPFFKE 173

Qy        168 VEIQLGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEY 227
```

```
Db        174 FVINLGYFFVLFGAFVIVGAGNAVNLTDGLDGLAIVPVMIAAATFGVIAYLAGNAVFANY 233

Qy        228 LLIPNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQE 287

Db        234 LQINFVPGTGELAVIVGAVIGAGLGFLWFNAPPAAIFMGDTGSLALGGLIGSIAVATKHE 293

Qy        288 IVLFIMGGVFVMETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITV 347

Db        294 IVMVIVGGLFVMETLSVIIQVFWFKRTGRRVFLMAPIHHHFEKKGWTESQVVIRFWIISV 353

Qy        348 ILVLIGLATLKLR 360

Db        354 GLALLGLATLKLR 366


RESULT 15
G84955
phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13) [imported] - Buchnera sp
C;Species: Buchnera sp.
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: G84955
R;Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A;Reference number: A84930; MUID:20445173; PMID:10993077
A;Accession: G84955
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-357 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: mraY; BU219
C;Superfamily: phospho-N-acetylmuramoyl-pentapeptide-transferase
C;Keywords: transferase

  Query Match           49.8%;  Score 915.5;  DB 2;  Length 357;
  Best Local Similarity  46.2%;  Pred. No. 7.1e-62;
  Matches  166;  Conservative   86;  Mismatches  104;  Indels    3;  Gaps    1;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60

Db          1 MLIFFNKYL---HINLNILSYIPYRAIFSLLTSFFINLYIGPYFIYYFKKLQTYQIIRNN 57

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Db         58 GPKTHYSKKNTPTMGGIFIIFSILFSTILYCNLSNIYIWYVISILIGYGLIGFIDDYKKI 117

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Db        118 KYKNSQGLKLKWKYFFLSIIAFIFICMIKINNKDIISTELIIPFCIKNDFEINYLYIFLS 177

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Db        178 YFVLVGTSNAVNLTDGLDGLAIMPVIFLTCGLTLISLFSDNINISHYLHVQYVKNSTELA 237

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Db        238 ILCMAIVGSGLGFLWFNSYPAKVFMGDVGSLALGGSLGAIAILLHQELLLIIMGGIFVFE 297

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKL 359

Db        298 TISVILQIISFKIRKKRIFQMAPVHHHYEVKGILEPLIIVRFWIVSLILLLISLISLKV 356


Search completed: May  7, 2004, 08:40:55
Job time : 22 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 08:40:38 ; Search time 48 Seconds
(without alignments)
2081.751 Million cell updates/sec

Title: US-10-089-787-2
Perfect score: 1837
Sequence: 1 MLLLLAEYLQQFYKGFGVFQ.....RFWIITVILVLIGLATLKLR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1824	99.3	360	9	US-09-815-242-12011	Sequence 12011, A
2	1824	99.3	360	12	US-10-282-122A-66653	Sequence 66653, A
3	1689	91.9	360	12	US-10-282-122A-67901	Sequence 67901, A
4	1658	90.3	360	12	US-10-282-122A-69713	Sequence 69713, A
5	1292	70.3	360	12	US-10-282-122A-77459	Sequence 77459, A
6	1274	69.4	360	12	US-10-282-122A-66896	Sequence 66896, A
7	1268	69.0	360	9	US-09-741-669-451	Sequence 451, App
8	1268	69.0	360	9	US-09-815-242-10028	Sequence 10028, A
9	1268	69.0	360	12	US-10-282-122A-56412	Sequence 56412, A
10	1265	68.9	360	12	US-10-282-122A-77920	Sequence 77920, A
11	1264	68.8	360	12	US-10-282-122A-76018	Sequence 76018, A
12	1261	68.6	360	12	US-10-282-122A-75186	Sequence 75186, A
13	1258	68.5	360	9	US-09-815-242-14075	Sequence 14075, A
14	1257	68.4	360	9	US-09-815-242-11691	Sequence 11691, A
15	1257	68.4	360	12	US-10-282-122A-55841	Sequence 55841, A
16	1257	68.4	360	12	US-10-282-122A-59586	Sequence 59586, A
17	1250	68.0	360	12	US-10-282-122A-68814	Sequence 68814, A
18	1228	66.8	352	12	US-10-282-122A-73233	Sequence 73233, A
19	1216	66.2	360	9	US-09-815-242-11186	Sequence 11186, A
20	1216	66.2	360	12	US-10-282-122A-58419	Sequence 58419, A
21	1205.5	65.6	389	12	US-10-282-122A-50013	Sequence 50013, A
22	1202.5	65.5	389	12	US-10-282-122A-51377	Sequence 51377, A
23	1185.5	64.5	389	12	US-10-282-122A-49105	Sequence 49105, A
24	1118	60.9	372	12	US-10-282-122A-44716	Sequence 44716, A
25	1099	59.8	360	12	US-10-282-122A-66064	Sequence 66064, A
26	1094	59.6	360	12	US-10-282-122A-65130	Sequence 65130, A
27	972.5	52.9	365	12	US-10-282-122A-63058	Sequence 63058, A
28	901.5	49.1	239	12	US-10-282-122A-61467	Sequence 61467, A
29	812.5	44.2	351	12	US-10-282-122A-47085	Sequence 47085, A
30	782.5	42.6	353	12	US-10-282-122A-54319	Sequence 54319, A
31	755.5	41.1	215	12	US-10-282-122A-48088	Sequence 48088, A
32	709.5	38.6	353	12	US-10-335-977-5598	Sequence 5598, Ap
33	709.5	38.6	376	12	US-10-335-977-5599	Sequence 5599, Ap
34	708.5	38.6	353	9	US-09-815-242-11328	Sequence 11328, A
35	708.5	38.6	353	12	US-10-282-122A-58718	Sequence 58718, A
36	671	36.5	324	12	US-10-282-122A-46783	Sequence 46783, A
37	652	35.5	321	9	US-09-815-242-12347	Sequence 12347, A
38	652	35.5	321	9	US-09-815-242-12794	Sequence 12794, A
39	652	35.5	321	12	US-10-282-122A-44257	Sequence 44257, A
40	642	34.9	317	12	US-10-282-122A-52288	Sequence 52288, A
41	639	34.8	324	12	US-10-282-122A-60973	Sequence 60973, A
42	637	34.7	422	12	US-10-282-122A-48543	Sequence 48543, A
43	636	34.6	321	12	US-10-282-122A-70722	Sequence 70722, A
44	629.5	34.3	321	12	US-10-282-122A-71642	Sequence 71642, A
45	617.5	33.6	357	14	US-10-156-761-13646	Sequence 13646, A

ALIGNMENTS

RESULT 1
US-09-815-242-12011
; Sequence 12011, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12011
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa

US-09-815-242-12011

Query Match 99.3%; Score 1824; DB 9; Length 360;
Best Local Similarity 99.4%; Pred. No. 6.9e-166;
Matches 358; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 2
US-10-282-122A-66653
; Sequence 66653, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66653
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66653

Query Match 99.3%; Score 1824; DB 12; Length 360;
Best Local Similarity 99.4%; Pred. No. 6.9e-166;
Matches 358; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||||||||||||||||||||||||||||:||||||||||||||||||||||||||||
Db        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 3
US-10-282-122A-67901
; Sequence 67901, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67901
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-10-282-122A-67901

  Query Match             91.9%;  Score 1689;  DB 12;  Length 360;
  Best Local Similarity   90.3%;  Pred. No. 5.6e-153;
  Matches  325;  Conservative   18;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||:||| |||||:||||| |||||||:|||||||||||| |||||||||
Db          1 MLLLLAEYLQQFHKGFAVFQYLSLRGILGVLTALSLALWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||||||| ||||||||||:|||:||||||||||||||||:||| |||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSAIAVSTLLWADLSNRYVWVVLIVTLAFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||: |||||| || | :|||||:| :| | | ||: |||||
Db        121 IEKNSRGLPSRWKYFWQSVFGLAAAVFLYKTAPTSVETTLILPFIKDVTIPLGVGFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||| |||||||||||||||||||| |||:||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGSGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||| ||:|||||||||||||||||||||||||||||||||||||||||||||:||:|
Db        241 VFCGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGIFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:||||||||:|||||||||||||||:|||||||||||||||||||||||||||||
Db        301 TLSVVIQVASFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360


RESULT 4
US-10-282-122A-69713
; Sequence 69713, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 69713
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-10-282-122A-69713

  Query Match             90.3%;  Score 1658;  DB 12;  Length 360;
  Best Local Similarity   89.7%;  Pred. No. 5.1e-150;
  Matches  323;  Conservative   17;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |||||||:|||||||| |||||:||||| ||||:||| ||||||||||: ||||:||||
Db          1 MLLLLAEFLQQFYKGFAVFQYLSLRGILGVLTALTLSLCLGPWMIRTLQMRQIGQSVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||||||| ||||||||||::| ||||||||||||||||:|| ||||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSSIGISTLLWADLSNRYVWVVLLVTFLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||: ||:||| || :  |||||||||| | | ||| |:|||
Db        121 IEKNSRGLPSRWKYFWQSVFGLCAAIFLYTTAPSATETTLIVPMLKDVRIPLGIGFIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||| |||||||||||||||||||| ||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||  ||:|||||||||||||||||||||||||||||||||:|||||||:|||||||||||
Db        241 VFSGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTMAVIVRQEMVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:|||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        301 TLSVVIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360


RESULT 5
US-10-282-122A-77459
; Sequence 77459, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
```

```
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77459
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Vibrio cholerae
US-10-282-122A-77459

  Query Match             70.3%;  Score 1292;  DB 12;  Length 360;
  Best Local Similarity   66.9%;  Pred. No. 5.2e-115;
  Matches  241;  Conservative   51;  Mismatches   68;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |:: ||| || ::  | :|:||: | |:|:|||| :|||:|| ||: ||: |||| |||:
Db          1 MIIWLAELLQPYFSFFRLFEYLSFRAIVSILTALGISLWMGPRMIKRLQMLQIGQVVRNE 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || |: |||||:| ||| || | | :||:|:|||||||
Db         61 GPESHFSKRGTPTMGGVMILAAITITVLLWADLTNPYVWAVLAVLLGYGAVGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||: || :|||||||    :  |  ||   :    | |:||  | |  |||: ::|||
Db        121 VRKNTDGLIARWKYFWQSAIALVVAFALYAHGQDTAATQLVVPFFKDVMPQLGLMYIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:||||||||||||||||||:||       : :||| || || || :| : ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAAGFAAIAWATGNVNFANYLHIPYIPHSSELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              | | |:|||||||||||||||||||||||||||| |||||||:|||| || |||||||||
Db        241 VVCTAMVGAGLGFLWFNTYPAQVFMGDVGALALGGALGTIAVLVRQEFVLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:|| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISIVLVLIGLATLKVR 360


RESULT 6
US-10-282-122A-66896
; Sequence 66896, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66896
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Pasteurella multocida
US-10-282-122A-66896

  Query Match             69.4%;  Score 1274;  DB 12;  Length 360;
  Best Local Similarity   65.8%;  Pred. No. 2.7e-113;
  Matches  237;  Conservative   57;  Mismatches   66;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: | |:|||:| || |  |:|:| ||::|||| :|||:|| :|| ||: : || ||:|
Db          1 MLVWLGEFLQQYYSGFNVISYITVRAILALLTALLVSLWIGPKVIRRLQLLKFGQEVRHD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || |:||||:|:| |||  | | | :||||:||||||:
Db         61 GPESHFSKRGTPTMGGVMILFAITVSTLLWANLANPYVWFSLFVLLGYGAIGFVDDYRKI 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||: || :|||||| ||  : ||  :|    :    | |:||  | :  |||:|:::||
Db        121 TRKNTDGLIARWKYFWLSVIALVAAFGMYAIGKDTDATRLVVPFFKEIMPQLGLFYIILT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||||:||||  :  : :||| ||||| || :  : ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAGAFALIAWATGNVNFAEYLHIPYIKFSAELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||| |:|||||||||||||||||||||||:|||| ||| :||:|||| :| |||||||:|
Db        241 VFCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGVVAVLVRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:||  :|:|||||||||||||:|||||||||:::|||:|| |||||
Db        301 TLSVILQVGSYKLRKQRIFRMAPIHHHFELKGWPEPRVIVRFWIISLMLVLVGLVTLKLR 360


RESULT 7
US-09-741-669-451
; Sequence 451, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  TITLE OF INVENTION:  Genes identified as required for
;  TITLE OF INVENTION:  proliferation of E. coli
;  FILE REFERENCE:  ELITRA.009A
;  CURRENT APPLICATION NUMBER: US/09/741,669
```

```
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER:   US 60/173005
;  PRIOR FILING DATE:   1999-12-23
;  NUMBER OF SEQ ID NOS:  481
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 451
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-741-669-451

  Query Match             69.0%;  Score 1268;  DB 9;  Length 360;
  Best Local Similarity   66.4%;  Pred. No. 1e-112;
  Matches  239;  Conservative   51;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :||||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||| |||   :  : :||: || || ||  :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| ||||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 8
US-09-815-242-10028
; Sequence 10028, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10028
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-09-815-242-10028

  Query Match             69.0%;  Score 1268;  DB 9;  Length 360;
  Best Local Similarity   66.4%;  Pred. No. 1e-112;
  Matches  239;  Conservative   51;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :||||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||| |||   :  : :||: || || ||  :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| ||||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 9
US-10-282-122A-56412
; Sequence 56412, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
```

; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56412
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-56412

Query Match 69.0%; Score 1268; DB 12; Length 360;
Best Local Similarity 66.4%; Pred. No. 1e-112;
Matches 239; Conservative 51; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||     || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :|||| || |||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 10
US-10-282-122A-77920
; Sequence 77920, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77920
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Yersinia pestis
US-10-282-122A-77920

Query Match 68.9%; Score 1265; DB 12; Length 360;
Best Local Similarity 65.6%; Pred. No. 2e-112;
Matches 236; Conservative 54; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: ||||| :|| || || ||| | |:|:|||| :|||:|| :|  ||  |||| ||||
Db          1 MLVWLAEYLVKFYSGFNVFSYLTFRAIVSLLTALFISLWMGPHLIAWLQKLQIGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :|| :| || |:||  || ||| || : : :| :|::||||||
Db         61 GPESHFSKRGTPTMGGLMILFSITISVLMWAYPSNPYVWCVLFILIGYGIVGFIDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||::|| :||||||||:  : ||  :|    :    | |:||  | :  |||: :|:|
Db        121 VRKNTKGLIARWKYFWQSIIALAAAFTMYSIGKDTSATELVVPFFKDIMPQLGLLYVLLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||| ||    :  : :||| || || || :  ||||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNVNFAAYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| |||||||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 11
US-10-282-122A-76018
; Sequence 76018, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76018
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-10-282-122A-76018

Query Match 68.8%; Score 1264; DB 12; Length 360;
Best Local Similarity 66.1%; Pred. No. 2.5e-112;
Matches 238; Conservative 52; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||     || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :|||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|:
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| ||    :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||:||||||:|||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 12
US-10-282-122A-75186
; Sequence 75186, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75186
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Salmonella typhimurium
US-10-282-122A-75186

Query Match 68.6%; Score 1261; DB 12; Length 360;
Best Local Similarity 65.8%; Pred. No. 4.7e-112;
Matches 237; Conservative 53; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||     || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :|||:| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTSIVISVLLWAYPSNPYVWCVLVVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|:
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| ||    :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||:||||||:|||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 13
US-09-815-242-14075
; Sequence 14075, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14075
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(360)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-14075

Query Match 68.5%; Score 1258; DB 9; Length 360;
Best Local Similarity 65.8%; Pred. No. 9.2e-112;
Matches 237; Conservative 52; Mismatches 71; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| || || || |||| ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||||| :||||| || |||| || ||| |||| : :| ||:||||||| 
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| || :| | ||: : | |:|| | | |||:|:::|: 
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| ||  :  : :||: || || || :  ||||: 
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| ||||||||:| 
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:|||||||||||:::|||||||||:| 
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLXVR 360
```

RESULT 14
US-09-815-242-11691
; Sequence 11691, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11691
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-815-242-11691

Query Match 68.4%; Score 1257; DB 9; Length 360;
Best Local Similarity 65.6%; Pred. No. 1.1e-111;
Matches 236; Conservative 52; Mismatches 72; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| || || || |||| ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLAFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||||| :||||| :| |||| || ||| || | : :| ||:||||||| 
Db         61 GPESHFSKRGTPTMGGIMILTAITVSVLLWAYPSNPYVWCVLTVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| || :| | ||: : | |:|| | | |||:|:::| 
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATELVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| ||  :  : :||: || || || :  ||||: 
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| ||||||||:| 
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:|||||||||||:::|||||||||:| 
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 15
US-10-282-122A-55841
; Sequence 55841, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

```
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 55841
;   LENGTH: 360
;   TYPE: PRT
;   ORGANISM: Enterobacter cloacae
US-10-282-122A-55841

  Query Match             68.4%;  Score 1257;  DB 12;  Length 360;
  Best Local Similarity   65.6%;  Pred. No. 1.1e-111;
  Matches  236;  Conservative   52;  Mismatches   72;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| :|||| :| |||| || ||| || | : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVVSVLLWAYPSNPYVWCVLTVLVGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATELVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| ||    :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::||||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


Search completed: May  7, 2004, 08:46:23
Job time : 49 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 08:37:38 ; Search time 23 Seconds
(without alignments)
808.059 Million cell updates/sec

Title: US-10-089-787-2
Perfect score: 1837
Sequence: 1 MLLLLAEYLQQFYKGFGVFQ.....RFWIITVILVLIGLATLKLR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1824	99.3	467	4	US-09-252-991A-24194	Sequence 24194, A
2	1250	68.0	360	4	US-09-543-681A-7051	Sequence 7051, Ap
3	1118	60.9	372	4	US-09-328-352-6169	Sequence 6169, Ap
4	1038.5	56.5	343	4	US-09-489-039A-12439	Sequence 12439, A
5	970.5	52.8	368	4	US-09-540-236-3138	Sequence 3138, Ap
6	636	34.6	334	4	US-09-134-001C-3851	Sequence 3851, Ap
7	593	32.3	331	4	US-09-107-532A-5276	Sequence 5276, Ap
8	583.5	31.8	327	4	US-09-134-000C-4685	Sequence 4685, Ap
9	546.5	29.7	349	4	US-09-198-452A-973	Sequence 973, App
10	520	28.3	326	3	US-08-986-768-2	Sequence 2, Appli
11	245	13.3	386	4	US-09-107-532A-4764	Sequence 4764, Ap
12	231.5	12.6	106	4	US-09-634-238-311	Sequence 311, App
13	222.5	12.1	375	4	US-09-134-001C-5489	Sequence 5489, Ap
14	222	12.1	69	4	US-09-489-039A-12452	Sequence 12452, A
15	192	10.5	341	2	US-08-846-762-92	Sequence 92, Appl
16	186.5	10.2	268	4	US-09-134-000C-4101	Sequence 4101, Ap
17	177	9.6	359	4	US-09-252-991A-28730	Sequence 28730, A
18	176	9.6	403	4	US-08-311-731A-157	Sequence 157, App
19	175	9.5	303	2	US-08-846-762-20	Sequence 20, Appl
20	175	9.5	303	2	US-08-846-762-91	Sequence 91, Appl
21	155.5	8.5	355	2	US-08-846-762-93	Sequence 93, Appl
22	151	8.2	373	4	US-09-543-681A-4417	Sequence 4417, Ap
23	150.5	8.2	393	4	US-09-489-039A-7988	Sequence 7988, Ap
24	121	6.6	85	4	US-08-936-165A-423	Sequence 423, App
25	119	6.5	295	4	US-09-543-681A-5660	Sequence 5660, Ap
26	117.5	6.4	452	4	US-09-489-039A-12558	Sequence 12558, A
27	117	6.4	426	4	US-09-489-039A-11995	Sequence 11995, A
28	117	6.4	466	4	US-09-328-352-4381	Sequence 4381, Ap
29	116.5	6.3	471	4	US-09-711-164-444	Sequence 444, App
30	113.5	6.2	387	4	US-09-540-236-3550	Sequence 3550, Ap
31	113	6.2	398	4	US-09-134-001C-4353	Sequence 4353, Ap
32	112.5	6.1	104	4	US-09-634-238-312	Sequence 312, App
33	111	6.0	255	2	US-09-154-802-3	Sequence 3, Appli
34	111	6.0	255	3	US-09-373-029-3	Sequence 3, Appli
35	110.5	6.0	457	4	US-09-489-039A-8293	Sequence 8293, Ap
36	110	6.0	400	4	US-09-543-681A-5821	Sequence 5821, Ap
37	109.5	6.0	540	4	US-09-252-991A-18967	Sequence 18967, A
38	109.5	6.0	767	4	US-09-540-236-2346	Sequence 2346, Ap
39	108.5	5.9	461	4	US-09-107-532A-5964	Sequence 5964, Ap
40	106.5	5.8	491	4	US-09-543-681A-8251	Sequence 8251, Ap
41	105.5	5.7	416	3	US-09-333-208-2	Sequence 2, Appli
42	105.5	5.7	416	3	US-09-333-254-2	Sequence 2, Appli
43	105.5	5.7	416	4	US-09-183-270-2	Sequence 2, Appli
44	105.5	5.7	510	4	US-09-489-039A-9866	Sequence 9866, Ap
45	105.5	5.7	548	4	US-09-543-681A-4994	Sequence 4994, Ap

ALIGNMENTS

RESULT 1
US-09-252-991A-24194
; Sequence 24194, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24194
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24194

Query Match 99.3%; Score 1824; DB 4; Length 467;
Best Local Similarity 99.4%; Pred. No. 2.2e-197;
Matches 358; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        108 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIRQIGQAVRND 167

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        168 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 227

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        228 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 287

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        288 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 347

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        348 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 407

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              |||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        408 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 467
```

RESULT 2
US-09-543-681A-7051
; Sequence 7051, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7051
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7051

Query Match 68.0%; Score 1250; DB 4; Length 360;
Best Local Similarity 65.0%; Pred. No. 1.1e-132;
Matches 234; Conservative 56; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: ||||| :::  | || ||| | |: :|||| ::||:|| :|  ||  |||| |||:
Db          1 MLVWLAEYLVKYHTFFNVFSYLTFRAIVGLLTALIIALWMGPHLIAWLQKMQIGQVVRNE 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| :||:|||||  | | ||| ||:| : :| ||::||||||
Db         61 GPESHFSKRGTPTMGGIMILFSIAVSTLLWARLDNPYVWCVLLVLIGYGIIGFIDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |::||| :|||||||||  :  |  :|   :    | |:||  | |  |||: :::|
Db        121 VRKDTRGLIARWKYFWQSVLALAVAFSMYAIGKDTPATQLVVPFFKDVMPQLGMLYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:||||||||||||||||||  ||   :  : :||| ||  ||  ||  :  ||||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNVNFASYLKIPYLMHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:||||||||||||||||||||||:|||| |||||||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:|||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 3
US-09-328-352-6169
; Sequence 6169, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6169
; LENGTH: 372
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6169

Query Match 60.9%; Score 1118; DB 4; Length 372;
Best Local Similarity 58.6%; Pred. No. 9.9e-118;
Matches 218; Conservative 62; Mismatches 80; Indels 12; Gaps 3;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||| | | |  ::  | | :|||||  :|||||:|:: | ||| ||| ||  : ||||  :
Db          1 MLLWLFEQLAGYHSSFQVVRYLTLRSLLSVLTSLTIGLVLGPIMIRKLQALKYGQAVSSF 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |::|  | ||||||| ||| :| |||||||||| |||:|| | ::|||:|| ||: |:
Db         61 APENHAKKMGTPTMGGILILLSIGISTLLWADLSNPYVWIVLGVMVVPGAVGWADDWIKI 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYM--TAETPIETT-----LIVPMLKSVEI--- 170
              |:: |||:| |:|| ||  :|| : ||:  | ::  | |       |::|  |:: |
Db        121 RYKDNAGLPARKKFFWTSVASLGAGIALYLIATQQSNAEYTANMLDLLIPFFKNLSIPLS 180

Qy        171 --QLGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYL 228
                ||: |:| || || |:|||||||||||||||||| ||||  ||:| ||||:::|| ||
Db        181 IVPLGLAFIVFTYLVINGASNAVNLTDGLDGLAIMPVVMVATGLGVFAYLSGDIRFANYL 240

Qy        229 LIPNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEI 288
              || |    ||:| |:|::|||| |||:| :|||||||||||||||| ||||||:|||||
Db        241 HIPYVKYTSELVVICSAMIGAGLAFLWYNAHPAQVFMGDVGALALGAMLGTIAVMVRQEI 300

Qy        289 VLFIMGGVFVMETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVI 348
              |  ||||||||| :|| :|: | ::  :||| |||:|||:| :|| : :|::|||||::
Db        301 VFAIMGGVFVMEAVSVFLQIGSLRMRNKRVFLMAPLHHHYEKQGWKETQVVIRFWIITIM 360

Qy        349 LVLIGLATLKLR 360
              ||::|| |||||
Db        361 LVVLGLMTLKLR 372
```

RESULT 4
US-09-489-039A-12439
; Sequence 12439, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12439
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12439

Query Match 56.5%; Score 1038.5; DB 4; Length 343;
Best Local Similarity 61.7%; Pred. No. 8.5e-109;
Matches 201; Conservative 45; Mismatches 75; Indels 5; Gaps 1;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db         16 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLAFGQVVRND 75

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|||||:| |||  || ||| || | : :| ||:|||||||
Db         76 GPESHFSKRGTPTMGGIMILTAITVSVLLWAYPSNPYVWCVLTVLIGYGIIGFVDDYRKV 135

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        136 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATELVVPFFKDVMPQLGLFYILLA 195

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||  ||    :  : :||: || || || :  ||||:
Db        196 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 255

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        256 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 315

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHH 326
```

```
Db        316 TLSVILQVGSFKLR-----RSAHLPH 336


RESULT 5
US-09-540-236-3138
; Sequence 3138, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 2709.2005-001
;  CURRENT APPLICATION NUMBER: US/09/540,236
;  CURRENT FILING DATE: 2000-04-04
;  NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 3138
;   LENGTH: 368
;   TYPE: PRT
;   ORGANISM: M.catarrhalis
US-09-540-236-3138

  Query Match            52.8%;  Score 970.5;  DB 4;  Length 368;
  Best Local Similarity  52.6%;  Pred. No. 4.6e-101;
  Matches  184;  Conservative   69;  Mismatches   86;  Indels   11;  Gaps    2;

Qy         22 LTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPTMGGALILT 81
Db         19 LTLRALLAVITSLLIVIIGGKPVIHYLRTLKYGQAVRDDGPKTHLAKQGTPTMGGVLILV 78

Qy         82 AIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWKYFWQSVFG 141
Db         79 AIGIATLAWADLGNPYVWILMVVMVIFGAVGWADDWLKIKHKNPQGLIARKKYFWLSVGS 138

Qy        142 IGAAVFLYMTAETPIETTLI------VPMLKSV-----EIQLGIFFVVLTYFVIVGSSNA 190
Db        139 LFAGCSMYYIATQQDTNTMIAMQDMLIPLFKDIVIPFSAIPLGLGFIIATYFVLAGSSNA 198

Qy        191 VNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELIVFCAALVGAG 250
Db        199 VNLTDGLDGLVILPVVLVAAGLGVFAYISGSPHFADYMHVPYIAYNAEVTIVCAAMIGAG 258

Qy        251 LGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVMETLSVMIQVAS 310
Db        259 LGFLWYNAAPADVFMGDVGALSLGGMLGTIAVMTRQELAFAIMGGIFVAEAVSVILQVGS 318

Qy        311 FKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
Db        319 YRLRKQRIFLMAPLHHHFEELGLKETKVVARFYIVCIILVVLGLMTLKLR 368


RESULT 6
US-09-134-001C-3851
; Sequence 3851, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3851
;   LENGTH: 334
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3851

  Query Match            34.6%;  Score 636;  DB 4;  Length 334;
  Best Local Similarity  42.3%;  Pred. No. 2.5e-63;
  Matches  140;  Conservative   70;  Mismatches   93;  Indels   28;  Gaps    7;

Qy         31 LTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPTMGGALILTAIAIST--- 87
Db         20 IIALLITFILVPILIPTLKRMKFGQSIREEGPQSHMKKTGTPTMGGLTFLISIIITSIIA 79

Qy         88 LLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWKYFWQSVFGIGAAVF 147
Db         80 IIFVDHSNPII-LLLFVTIGFGLIGFIDDYIIVVKKNNQGLTSKQKFLAQIII----AVI 134

Qy        148 LYMTAE----TPIETTLIVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIM 203
Db        135 FFVLSDVFHLVHFTTDLHIPFV-NFDIPLSFAYVIFIVFWQVGFSNAVNLTDGLDGLAT- 192

Qy        204 PTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFLWFNTYPAQV 263
Db        193 -------GLSI-------IGFAMYAVMSYMLDSPAIGIFCIIMIFALLGFLPYNLNPAKV 238

Qy        264 FMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVMETLSVMIQVASFKLTGRRVFRMAP 323
Db        239 FMGDTGSLALGGIFATISIMLNQELSLILIGFVFVVETLSVMLQVASYKLTKKRIFKMSP 298

Qy        324 IHHHFELKGWPDPRVIVRFWIITVILVLIGL 354
Db        299 IHHHFELSGWGEWKVVTVFWTVGLITGLIGL 329


RESULT 7
US-09-107-532A-5276
; Sequence 5276, Application US/09107532A
; Patent No. 6583275
;  GENERAL INFORMATION:
;       APPLICANT: Lynn A Doucette-Stamm  and David Bush
;       TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
;                           ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
;       NUMBER OF SEQUENCES: 7310
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: GENOME THERAPEUTICS CORPORATION
;            STREET: 100 Beaver Street
;            CITY: Waltham
;            STATE: Massachusetts
;            COUNTRY: USA
;            ZIP: 02354
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: CD/ROM ISO9660
;            COMPUTER: PC
;            OPERATING SYSTEM: <Unknown>
;            SOFTWARE: ASCII
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/107,532A
;            FILING DATE: 30-Jun-1998
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 60/085,598
;            FILING DATE: 14 May 1998
;            APPLICATION NUMBER: 60/051571
;            FILING DATE: July 2, 1997
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Ariniello, Pamela Deneke
;            REGISTRATION NUMBER: 40,489
;            REFERENCE/DOCKET NUMBER: GTC-012
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (781)893-5007
;            TELEFAX: (781)893-8277
;  INFORMATION FOR SEQ ID NO: 5276:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 331 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       HYPOTHETICAL: YES
;       ORIGINAL SOURCE:
```

; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...331
; SEQUENCE DESCRIPTION: SEQ ID NO: 5276:
US-09-107-532A-5276

Query Match 32.3%; Score 593; DB 4; Length 331;
Best Local Similarity 38.0%; Pred. No. 1.8e-58;
Matches 128; Conservative 75; Mismatches 102; Indels 32; Gaps 7;

Qy 25 RGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPTMGGAL-----I 79
Db 16 QALIPIVSSCAMTIAAMPLFIGYFQMKKQGQAIREEGPKWHNSKAGTPTMGGLVFLIGSI 75

Qy 80 LTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWKYFWQSV 139
Db 76 LTGIWVGA--WQKQLTPTLFILLFVLALYGVIGFLDDFIKIFKKRNMGLNSKQKLIGQIL 133

Qy 140 FGIGAAVF--LYMTAETPIETTLIVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLTDGL 197
Db 134 GGI---IFYLVYRSEGYPGVLNFF-----GLDLPLGIIYGIFALFWLVGFSNAVNLTDGI 185

Qy 198 DGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFLWFN 257
Db 186 DGL-------VAGL--------GSISFAAYGLIAWHQHQYDVLIICLSVLGGLLGFFVYN 230

Qy 258 TYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVMETLSVMIQVASFKLTGRR 317
Db 231 RKPAKIFMGDVGSLALGGLLAAISIMLNQEWTLLLIGLIYVMETASVMLQVTSFKLTGKR 290

Qy 318 VFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGL 354
Db 291 IFKMSPIHHHFEMCGWSEWKIDTVFWLISIVTSLITL 327

RESULT 8
US-09-134-000C-4685
; Sequence 4685, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4685
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4685

Query Match 31.8%; Score 583.5; DB 4; Length 327;
Best Local Similarity 38.6%; Pred. No. 2.1e-57;
Matches 125; Conservative 68; Mismatches 108; Indels 23; Gaps 6;

Qy 28 LSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGPQSHLSKKGTPTMGGALILTAIAIST 87
Db 14 IPIVVSFAITVSVMPLFIGYFQMKKQGQVTREDGPTWHSVKTGTPTMGGVVFLVASLITS 73

Qy 88 LLWADLSNRY---VWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWKYFWQSVFGIGA 144
Db 74 LAMGLFFHQFTPSLLIILFILVLYGLLGYLDDFIKVFKKRNMGLNSRQKLIGQ-IF--GG 130

Qy 145 AVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIMP 204
Db 131 LVFYFVYRSEGFSDTL--DLFGVAEVPLGIFYGVFIIFWLVGFSNAVNLTDGIDGL---- 184

Qy 205 TVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVF 264
Db 185 ---VAGL--------GTISFGTYAIIAWKQQQFDVVIICLSVIGGLIGFFPYNRKPAKIF 233

Qy 265 MGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVMETLSVMIQVASFKLTGRRVFRMAPI 324
Db 234 MGDVGSLALGGLLAAISIILHQEWTLLLIGLVYVCETASVILQVASFKLFGRRIFKMSPI 293

Qy 325 HHHFELKGWPDPRVIVRFWIITVI 348
Db 294 HHHFEMCGWSEWKIDFVFWSVGLI 317

RESULT 9
US-09-198-452A-973
; Sequence 973, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 973
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-973

Query Match 29.7%; Score 546.5; DB 4; Length 349;
Best Local Similarity 35.1%; Pred. No. 3.5e-53;
Matches 129; Conservative 75; Mismatches 137; Indels 27; Gaps 9;

Qy 1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
Db 1 MIPLIPMFLKQ-----SLFFSLALTGMTTLVLTVSLGVPVMKWLKRKNYRDYIHKEYCEK 55

Qy 61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
Db 56 LEMLHKDKAEVPTGGGVLLFISLIASLLVWLPWGKFSTWFFIILLTCYAGLGWYDDRIKI 115

Qy 121 IEKNSRGLPSRWKYFWQ---SVFGIGAAVFLYMTAETPIETTLIVPMLK---SVEIQLG- 173
Db 116 KRKQGHGLKAKHKFMVQIAIAAFTLIALPYIYGSTE-PL-WTLKIPFMEGMLSLPFWLGK 173

Qy 174 IFFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALG-IFCYLSGNVKFAEYLLIPN 232
Db 174 VFCLGLALVAIIGTSNAVNLTDGLDGLA--AGTMSFAALGFIFVALRSS----------T 221

Qy 233 VPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFI 292
Db 222 IPIAQDVAYVLAALVGACIGFLWYNGFPAQLFMGDTGSLLLGGLLGSCAVMLRAECILVV 281

Qy 293 MGGVFVMETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLI 352
Db 282 IGGVFVAEAGSVILQVLSCRLRKKRLFLCSPLHHHYEYQGLPETKIVMRFWIFSFVCAGL 341

Qy 353 GLATLKLR 360
Db 342 GIAAVLWR 349

RESULT 10
US-08-986-768-2
; Sequence 2, Application US/08986768
; Patent No. 6271000
; GENERAL INFORMATION:
; APPLICANT: Rosteck Jr, Paul R
; TITLE OF INVENTION: Streptococcus Pneumoniae Gene Sequence
; TITLE OF INVENTION: MraY

; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: US
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/986,768
; FILING DATE: December 8, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: X-11753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317/276-3334
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 326 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-986-768-2

Query Match 28.3%; Score 520; DB 3; Length 326;
Best Local Similarity 36.9%; Pred. No. 3.1e-50;
Matches 129; Conservative 60; Mismatches 111; Indels 50; Gaps 9;

```
Qy          22 LTLRGILSVLTALSLSLWLGPWMIRTLQIPQI-GQAVRNDGPQSHLSKKGTPTMGGALIL 80
               ||| ||                 |  |:  :  || || :  |  | | :| |||||||| : |
Db          14 LTLVGI--------------PAFIQFYRKAQITGQQMHEDVKQ-HQAKAGTPTMGGLVFL 58

Qy          81 TAIAISTLLWADLSNRY---VWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWKYFWQ 137
                  :   :|  |:::   | ::| : :|:| :|::||: ||  | : ||  : |   |
Db          59 ITSVLVAPFFALFSSQFSNNVGMILFILVLYGLVGFLDDFLKVFRKINEGLNPKQKLALQ 118

Qy         138 SVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLTDGL 197
               : |:   :|:          |  :    :|  |:: |  | :|| |||||||||:
Db         119 LLGGV--IFYLFYERGGDILSVFGYP------VHLGFFYIFFALFWLVGFSNAVNLTDGV 170

Qy         198 DGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFLWFN 257
               |||| :  |:   | |:  |: | :               ::::   |::|  |||  ||
Db         171 DGLASISVVISLFAYGVIAYVQGQM--------------DILLVILAMIGGLLGFFIFN 215

Qy         258 TYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVMETLSVMIQVASFKLT-GR 316
                 ||:|||||||:||||  |  |:: : ||  | |:| |:|  || |||:||: |||| |:
Db         216 HKPAKVFMGDVGSLALGGMLAAISMALHQEWTLLIIGIVYVFETTSVMMQVSYFKLTGGK 275

Qy         317 RVFRMAPIHHHFELKG-------WPDPRVIVRFWIITVILVLIGLATLKL 359
               |:||| |:|||||| |       | : :|   || : ::  |: || | |
Db         276 RIFRMTPVHHHFELGGLSGKGNPWSEWKVDFFFWGVGLLASLLTLAILYL 325
```

RESULT 11
US-09-107-532A-4764
; Sequence 4764, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4764:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 386 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...386
; SEQUENCE DESCRIPTION: SEQ ID NO: 4764:
US-09-107-532A-4764

Query Match 13.3%; Score 245; DB 4; Length 386;
Best Local Similarity 26.6%; Pred. No. 4.9e-19;
Matches 101; Conservative 65; Mismatches 131; Indels 82; Gaps 17;

```
Qy          22 LTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRNDGP-QSHLSKKGTPTMGGALIL 80
               : :|  |::|     ||| | | : : : : | |      | | :  ::|   |: ||  |
Db          15 MVIRFFLTIL----LSLILTP-VFKIISV-QTGMV---DKPNERRINKVPMPSAGGLPIF 65

Qy          81 TAIAISTLLW--ADLSNRYVWVVLVVTLLFGAIGWVDDYRKVIEKNSRGLPSRWKYFWQS 138
               :  ||||       :  |:  :|: :|:    | :||  ::  |              :|
Db          66 ISFVISTLFLFRKIIPQSYIVPILIASLVIIVTGVLDDKYELTPKQ-----------KS 113

Qy         139 VFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLTDGLD 198
               :  : |:: :|  |:  |: :  :| :    ||||      :| | | | :||||| ||||
Db         114 IGILIASLIIYFVADIRID-SFTLPFIG--YIQLGWLSFPVTIFWIFGITNAVNLIDGLD 170

Qy         199 GLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFLWFNT 258
               |||   :::    :|| |          |  : ||       :::||  || : :||  :|
Db         171 GLAAGISLIGLITIGIIGYF---FLHASTVYIP-------VVIFC--LVASIIGFFPYNF 218

Qy         259 YPAQVFMGDVGALALGAALGTIA------VIVRQEIVLFIMGGVFVMETLSVMIQVASFK 312
               |||::::|| ||| ||  :  ::       |     | | |: || | :|   :|:  :
Db         219 YPAKIYLGDTGALFLGFMMAVLSLQGLKNVTFVSSISLLIVMGVPVTDTFFAIIR----R 274

Qy         313 LTGRRVFRMAP---IHHHFELKGWPDPRVIVRFW-----------------IITVILV- 350
               |  |  |    :||      |:     ::  :                 | |:||:
Db         275 KANRVSFSTADKKHLHHRLISLGFTHKGAVLTIYAMALMFSFTAMVMNYTGRIGTIILII 334

Qy         351 -----------LIGLATLK 358
                         ||||   |
Db         335 AMLFAAILLFELIGLINEK 353
```

RESULT 12
US-09-634-238-311

; Sequence 311, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna C.
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul W.
; APPLICANT: Reid, Julian R.
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; TITLE OF INVENTION: them and methods for using them.
; FILE REFERENCE: 11000.1043U1
; CURRENT APPLICATION NUMBER: US/09/634,238
; CURRENT FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 311
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-634-238-311

Query Match 12.6%; Score 231.5; DB 4; Length 106;
Best Local Similarity 45.5%; Pred. No. 2.3e-18;
Matches 56; Conservative 14; Mismatches 36; Indels 17; Gaps 3;

```
Qy        159 TLIVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYL 218
              || ||:| :|    |:|:  | :|| ||||||||||||     |||
Db          1 TLWVPVLGNVNATW--FYVIFVVFWLVGFSNAVNLTDGLDGL-------VAG-------- 43

Qy        219 SGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALG 278
              : || | :|    |  :::: |   ||| |||| ||  |||:||||:|:||||  |
Db         44 QTTISFATYAVIAVREGRVDVLIICLVTVGAMLGFLMFNHKPAQIFMGDLGSLALGGMLA 103

Qy        279 TIA 281
              |
Db        104 VAA 106
```

RESULT 13
US-09-134-001C-5489
; Sequence 5489, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5489
; LENGTH: 375
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5489

Query Match 12.1%; Score 222.5; DB 4; Length 375;
Best Local Similarity 25.0%; Pred. No. 1.6e-16;
Matches 77; Conservative 51; Mismatches 103; Indels 77; Gaps 10;

```
Qy         69 KGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAI-----GWVDD------Y 117
              |   ||| :|| :  |   |   :          |  |: |||     | :||      |
Db         58 KPISVMGGTVILFSFLIGIWLGHPIERE------VKPLILGAITMYMVGLIDDIYDLRPY 111

Qy        118 RKVIEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAET------PIETTLIVPMLKSVEIQ 171
              |:  :                  | ||: :     |      ||  |          |
Db        112 LKLAGQ-----------------IVAALIVTFYGITIDFISLPIGPT----------IH 143

Qy        172 LGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIP 231
              ||| : :|   ||  :||:|| |||||||   : :    :|    |  |:
Db        144 FGIFSIPITVIWIVAITNAINLIDGLDGLASGVSALALMTIGFIAILQANI--------- 194

Qy        232 NVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVL- 290
                  :|: |  |:|: ||||::| :||::|:|| ||| :|  :| ::::  : |
Db        195 ------FIIMICCVLLGSLLGFLFYNFHPAKIFLGDSGALMIGFIIGFLSLLGFKNITFI 248

Qy        291 -----FIMGGVFVMETLSVMIQVASFKLTGRRVFR--MAPIHHHFELKGWPDPRVIVRFW 343
                  ::  |  ::||  ||:       |: : :   : :||      |:  : :: :
Db        249 ALFFPIVILAVPFIDTLFAMIRRMK---KGQHIMQADKSHLHHKLLALGYTHRQTVLLIY 305

Qy        344 IITVILVL 351
              | ::  |
Db        306 SIAIMFSL 313

RESULT 14
US-09-489-039A-12452
; Sequence 12452, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12452
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12452

Query Match 12.1%; Score 222; DB 4; Length 69;
Best Local Similarity 75.0%; Pred. No. 1.5e-17;
Matches 39; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

```
Qy        309 ASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              |     |:|:||||||||:||||||:||||||||||:::|||||||||:|
Db         18 APLSCAGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 69
```

RESULT 15
US-08-846-762-92
; Sequence 92, Application US/08846762A
; Patent No. 5994072
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: de Kievit, Teresa
; TITLE OF INVENTION: No. 5994072el Proteins Involved in the Synthesis and Assembly
; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa
; FILE REFERENCE: 6580-089
; CURRENT APPLICATION NUMBER: US/08/846,762A
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 92
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-08-846-762-92

Query Match 10.5%; Score 192; DB 2; Length 341;
```

```
 Best Local Similarity   24.7%;  Pred. No. 4e-13;
 Matches   83;  Conservative   57;  Mismatches  116;  Indels   80;  Gaps   16;

Qy          19 FQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRN---DGPQSHLSK-KGTPTM 74
               |  ||:  :|||     | |:                |::|   | | |  |    ||
Db           5 FFLLTIFFLLSVGLTYLLRLY---------------ALKNNIIDTPNSRSSHVTPTPRG 48

Qy          75 GGALILTAIAISTLLWADLSNRYVWVVLVVTLL-----FGAIGWVDDYRKVIEKNSRGLP 129
               ||  |: :  |  :|:  |   |: :: || |:        :|: ||:  :  :
Db          49 GGVAIVISFLIGIILFYFLG--YLPILSVVGLIVSGGVIALVGFWDDHGHIAAR------ 100

Qy         130 SRWKYFWQSVFGIGAAVF-LYMTAETPIETTLIVPMLKSVEIQLGIFFVVLTYFVIVGSS 188
                 |: :    || | |:     |:       :    |:|||  :     :|
Db         101 ------WRLLAHFSAAAFLLFCFGGFPVLN------VSGFIIELGIFGSLFGLLFLVWML 148

Qy         189 NAVNLTDGLDGLAIMPTVMVA-GALGIFCYLSG-NVKFAEYLLIPNVPGAGELIVFCAAL 246
               |  |  ||:||||   |     | : |: |:|| :::   :|::        :: |
Db         149 NLYNFMDGIDGLASAEAVTACIGMIAIY-YISGDHIELNSFLVL--------WLLACTV- 198

Qy         247 VGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVI-----VRQEIVLFIMGGVFVMET 301
                |||| :|  ||::|||| |:  ||  :|::|:       |      |: |:|:::
Db         199 ----LGFLLWNFPPAKIFMGDAGSGFLGLMIGSLAISAGWIDTRFFFCWLILLGLFIVDA 254

Qy         302 LSVMIQ--VASFK----------LTGRRVFRMAPI 324
                 :::  :  ||          :  ||  |  |:
Db         255 TWTLVRRVLGGFKVYEAHRSHGYQIASRRFKRHLPV 290


Search completed: May  7, 2004, 08:41:31
Job time : 24 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 7, 2004, 08:37:37 ; Search time 59 Seconds
(without alignments)
1724.019 Million cell updates/sec

Title: US-10-089-787-2
Perfect score: 1837
Sequence: 1 MLLLLAEYLQQFYKGFGVFQ.....RFWIITVILVLIGLATLKLR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1837	100.0	360	4	AAB73487	Aab73487 Pseudomon
2	1824	99.3	360	4	AAU36418	Aau36418 Pseudomon
3	1824	99.3	360	6	ABU38729	Abu38729 Protein e
4	1689	91.9	360	6	ABU39977	Abu39977 Protein e
5	1658	90.3	360	6	ABU41789	Abu41789 Protein e
6	1292	70.3	360	6	ABU49535	Abu49535 Protein e
7	1280	69.7	361	6	ABM68871	Abm68871 Photorhab
8	1274	69.4	360	6	ABU38972	Abu38972 Protein e
9	1268	69.0	360	4	AAU34435	Aau34435 E. coli c
10	1268	69.0	360	4	AAG98403	Aag98403 Escherich
11	1268	69.0	360	6	ABU28488	Abu28488 Protein e
12	1265	68.9	360	6	ABU49996	Abu49996 Protein e
13	1264	68.8	360	6	ABU48094	Abu48094 Protein e
14	1261	68.6	360	6	ABU47262	Abu47262 Protein e
15	1258	68.5	360	4	AAU38482	Aau38482 Salmonell
16	1257	68.4	360	4	AAU36098	Aau36098 Klebsiell
17	1257	68.4	360	6	ABU27917	Abu27917 Protein e
18	1257	68.4	360	6	ABU31662	Abu31662 Protein e
19	1250	68.0	360	6	ABU40890	Abu40890 Protein e
20	1228	66.8	352	6	ABU45309	Abu45309 Protein e
21	1216	66.2	360	4	AAU35593	Aau35593 Haemophil
22	1216	66.2	360	6	ABU30495	Abu30495 Protein e
23	1205.5	65.6	389	6	ABU22089	Abu22089 Protein e
24	1202.5	65.5	389	6	ABU23453	Abu23453 Protein e
25	1185.5	64.5	389	6	ABU21181	Abu21181 Protein e
26	1118	60.9	372	6	ABU16792	Abu16792 Protein e
27	1118	60.9	372	6	ADA34882	Ada34882 Acinetoba
28	1099	59.8	360	6	ABU38140	Abu38140 Protein e
29	1099	59.8	374	3	AAY74406	Aay74406 Neisseria
30	1094	59.6	360	6	ABP80568	Abp80568 N. gonorr
31	1094	59.6	360	6	ABU37206	Abu37206 Protein e
32	1069	58.2	376	3	AAY74404	Aay74404 Neisseria
33	972.5	52.9	365	6	ABU35134	Abu35134 Protein e
34	901.5	49.1	239	6	ABU33543	Abu33543 Protein e
35	893.5	48.6	1286	4	ABG25629	Abg25629 Novel hum
36	893.5	48.6	1286	4	ABG25994	Abg25994 Novel hum
37	893.5	48.6	1286	4	ABG25086	Abg25086 Novel hum
38	812.5	44.2	351	6	ABU19161	Abu19161 Protein e
39	782.5	42.6	353	6	ABU26395	Abu26395 Protein e
40	755.5	41.1	215	6	ABU20164	Abu20164 Protein e
41	709.5	38.6	376	2	AAW20932	Aaw20932 H. pylori
42	708.5	38.6	353	4	AAU35735	Aau35735 Helicobac
43	708.5	38.6	353	6	ABU30794	Abu30794 Protein e
44	682	37.1	376	3	AAY74405	Aay74405 Neisseria
45	671	36.5	324	6	ABU18859	Abu18859 Protein e

ALIGNMENTS

RESULT 1
AAB73487
ID AAB73487 standard; protein; 360 AA.
XX
AC AAB73487;
XX
DT 17-JUL-2001 (first entry)
XX
DE Pseudomonas aeruginosa MraY protein.
XX
KW MraY; phospho-N-acetylmuramoyl-pentapeptide translocase;
KW peptidoglycan biosynthesis; bacterial cell wall; infection;
KW drug screening; antibacterial; Pseudomonad; G+C rich bacterium.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200125251-A1.
XX
PD 12-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US027056.
XX
PR 04-OCT-1999; 99US-0157580P.
XX
PA (MERI) MERCK & CO INC.
XX
PI El-Sherbeini M, Azzolina B;
XX
DR WPI; 2001-308221/32.
DR N-PSDB; AAF86639.
XX
PT New MraY gene and enzyme of Pseudomonas aeruginosa, useful in vitro
PT assays for screening antibacterial compounds that target cell wall
PT biosynthesis, particularly for screening antibiotics against
PT Pseudomonads.
XX
PS Claim 15; Fig 1; 22pp; English.
XX
CC This sequence represents the Pseudomonas aeruginosa phospho-N-
CC acetylmuramoyl-pentapeptide translocase, designated MraY. The MraY
CC protein is involved in bacterial cell wall biosynthesis. It catalyses the
CC first step of the membrane cycle of peptidoglycan biosynthesis, i.e., the
CC transfer of an N-acetylmuramic acid peptide to a bactoprenol phosphate
CC carrier molecule. The MraY protein is useful in in vitro assays to screen
CC for antibacterial compounds that target cell wall biosynthesis.
CC Inhibitors of the MraY protein are useful in preventing the growth of
CC Pseudomonads and other G+C rich bacteria. Pseudomonas aeruginosa is an
CC opportunistic pathogen which causes infections in patients with burns,

CC neutropenia, or cystic fibrosis. Primers or probes derived from the mraY
CC gene are useful in nucleic acid amplification-based assays for detecting
CC the presence of a polynucleotide encoding Pseudomonas aeruginosa MraY
CC protein
XX
SQ Sequence 360 AA;

Query Match 100.0%; Score 1837; DB 4; Length 360;
Best Local Similarity 100.0%; Pred. No. 2.3e-193;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 2
AAU36418
ID AAU36418 standard; protein; 360 AA.
XX
AC AAU36418;
XX
DT 14-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa cellular proliferation protein #408.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS54277.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 12011; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 99.3%; Score 1824; DB 4; Length 360;
Best Local Similarity 99.4%; Pred. No. 6.1e-192;
Matches 358; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||
Db        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 3
ABU38729
ID ABU38729 standard; protein; 360 AA.
XX
AC ABU38729;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #24256.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX

PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA42599.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 66653; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 99.3%; Score 1824; DB 6; Length 360;
Best Local Similarity 99.4%; Pred. No. 6.1e-192;
Matches 358; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||
Db        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 4
ABU39977
ID ABU39977 standard; protein; 360 AA.
XX
AC ABU39977;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #25504.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pseudomonas putida.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA43847.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 67901; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 91.9%; Score 1689; DB 6; Length 360;
Best Local Similarity 90.3%; Pred. No. 4.6e-177;
Matches 325; Conservative 18; Mismatches 17; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||||||||||||:||| |||||:||||| |||||||:|||||||||||||| |||||||||
Db          1 MLLLLAEYLQQFHKGFAVFQYLSLRGILGVLTALSLALWLGPWMIRTLQIRQIGQAVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||||||| ||||||||||:|||:||||||||||||||||:||| ||||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSAIAVSTLLWADLSNRYVWVVLIVTLAFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||: |||||| || | :|||||:| :| | | ||: |||||
Db        121 IEKNSRGLPSRWKYFWQSVFGLAAAVFLYKTAPTSVETTLILPFIKDVTIPLGVGFVVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||||| |||||||||||||||||||| |||:||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGSGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||| ||:||||||||||||||||||||||||||||||||||||||||||||||||:||:|
Db        241 VFCGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGIFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:|||||||||:|||||||||||||||||:|||||||||||||||||||||||||||
Db        301 TLSVVIQVASFKLTGKRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 5
ABU41789
ID ABU41789 standard; protein; 360 AA.
XX
AC ABU41789;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #27316.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pseudomonas syringae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX

DR WPI; 2003-029926/02.
DR N-PSDB; ACA45659.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 69713; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 90.3%; Score 1658; DB 6; Length 360;
Best Local Similarity 89.7%; Pred. No. 1.2e-173;
Matches 323; Conservative 17; Mismatches 20; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |||||||:|||||||| |||||:||||| |||||:||| |||||||||: ||||:||||
Db          1 MLLLLAEFLQQFYKGFAVFQYLSLRGILGVLTALTLSLCLGPWMIRTLQMRQIGQSVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              |||||||| ||||||||||::| |||||||||||||||||:|| ||||||||||||||
Db         61 GPQSHLSKSGTPTMGGALILSSIGISTLLWADLSNRYVWVVLLVTFLFGAIGWVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||||||||||||||||||||: ||:||| || :  |||||||||| | |  ||| |:|||
Db        121 IEKNSRGLPSRWKYFWQSVFGLCAAIFLYTTAPSATETTLIVPMLKDVRIPLGIGFIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              |||||||||||||||||||||||||||||| |||||||||||||||||||| |||||||||
Db        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVGGALGIFCYLSGNVKFAEYLLIPYVPGAGELI 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||  ||:||||||||||||||||||||||||||||||||||||:|||||||:||||||||
Db        241 VFSGALIGAGLGFLWFNTYPAQVFMGDVGALALGAALGTMAVIVRQEMVLFIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||:|||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db        301 TLSVVIQVASFKLTGRRVFRMAPIHHHFELKGWPEPRVIVRFWIITVILVLIGLATLKLR 360
```

RESULT 6

ABU49535
ID ABU49535 standard; protein; 360 AA.
XX
AC ABU49535;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #35062.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Vibrio cholerae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA53405.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 77459; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 70.3%; Score 1292; DB 6; Length 360;
Best Local Similarity 66.9%; Pred. No. 2.6e-133;
Matches 241; Conservative 51; Mismatches 68; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              |:: ||| || ::  | :|:||: | |:|:|||| :|||:|| ||: ||: |||| |||:
Db          1 MIIWLAELLQPYFSFFRLFEYLSFRAIVSILTALGISLWMGPRMIKRLQMLQIGQVVRNE 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || |: |||||:| ||| || | | :||:|:|||||||
Db         61 GPESHFSKRGTPTMGGVMILAAITITVLLWADLTNPYVWAVLAVLLGYGAVGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||: || :|||||||   :  |  ||   :    | :||  | |  |||: ::|||
Db        121 VRKNTDGLIARWKYFWQSAIALVVAFALYAHGQDTAATQLVVPFFKDVMPQLGLMYIVLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||:||       : :||| || || || :| : ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAAGFAAIAWATGNVNPANYLHIPYIPHSSELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              | | |:|||||||||||||||||||||||||| |||||||:||| || |||||||||
Db        241 VVCTAMVGAGLGFLWFNTYPAQVFMGDVGALALGGALGTIAVLVRQEFVLVIMGGVFVME 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |:|| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSYKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISIVLVLIGLATLKVR 360
```

RESULT 7
ABM68871
ID ABM68871 standard; protein; 361 AA.
XX
AC ABM68871;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #1968.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 1968; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification

CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 361 AA;

Query Match 69.7%; Score 1280; DB 6; Length 361;
Best Local Similarity 66.4%; Pred. No. 5.5e-132;
Matches 239; Conservative 55; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||||  ::: || || ||| | |:|:||||:::||:|| ||  ||  |||| ||||
Db          1 MLVWLAEYLVKYHSGFNVFSYLTFRAIVSLLTALAIALWMGPRMIAFLQKLQIGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:||||||| ||| :| |||||  |:| ||| ||:| : :| :|:|||||||
Db         61 GPESHFSKRGTPTMGGLLILLSITISTLLWVRLNNPYVWCVLIVLIGYGIVGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||||| :| | :|  :    | |:||  | |  |||: :::||
Db        121 VRKDTKGLIARWKYFWQSVLALGVAFAMYSFGKDTPATQLVVPFFKDVMPQLGVLYILLT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:||||||||||||||||||| ||    :  : :||| || || || :  ||||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNVNFANYLHIPYLSHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| |||||||::||| :| ||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::||||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 8
ABU38972
ID ABU38972 standard; protein; 360 AA.
XX
AC ABU38972;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #24499.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pasteurella multocida.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA42842.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 66896; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 69.4%; Score 1274; DB 6; Length 360;
Best Local Similarity 65.8%; Pred. No. 2.5e-131;
Matches 237; Conservative 57; Mismatches 66; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: | |:|||:| || |  |:|:| ||::|||| :|||:|| :|| ||: : || ||:|
Db          1 MLVWLGEFLQQYYSGFNVISYITVRAILALLTALLVSLWIGPKVIRRLQLLKFGQEVRHD 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| || :|||||:|:| |||  | | | :||||:||||||:
Db         61 GPESHFSKRGTPTMGGVMILFAITVSTLLWANLANPYVWFSLFVLLGYGAIGFVDDYRKI 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              ||: || :|||||| ||  : ||  :|   :    | |:||  | :  |||:|:::||
Db        121 TRKNTDGLIARWKYFWLSVIALVAAFGMYAIGKDTDATRLVVPFFKEIMPQLGLFYIILT 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:|||||||||||||||||:||||  :  : :||| ||||| || :  : ||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVLVAGAFALIAWATGNVNFAEYLHIPYIKFSAELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              ||| |:|||||||||||||||||||:|||| ||| :||:|||| :| |||||||:|
Db        241 VFCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGVVAVLVRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
```

```
                ||||::|| |:||  :|:||||||||||||||||:||||||||||:::|||:|| |||||
Db        301 TLSVILQVGSYKLRKQRIFRMAPIHHHFELKGWPEPRVIVRFWIISLMLVLVGLVTLKLR 360


RESULT 9
AAU34435
ID   AAU34435 standard; protein; 360 AA.
XX
AC   AAU34435;
XX
DT   14-FEB-2002  (first entry)
XX
DE   E. coli cellular proliferation protein #16.
XX
KW   Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW   antibacterial; drug design.
XX
OS   Escherichia coli.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US009180.
XX
PR   21-MAR-2000; 2000US-0191078P.
PR   23-MAY-2000; 2000US-0206848P.
PR   26-MAY-2000; 2000US-0207727P.
PR   23-OCT-2000; 2000US-0242578P.
PR   27-NOV-2000; 2000US-0253625P.
PR   22-DEC-2000; 2000US-0257931P.
PR   16-FEB-2001; 2001US-0269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS52294.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids.
XX
PS   Example 3; SEQ ID NO 10028; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the genes,
CC   their use in the discovery of novel antibiotics, the essential genes
CC   themselves and the encoded proteins. The prokaryotes used are Escherichia
CC   coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC   Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC   useful for the identification of potential new targets for antibiotic
CC   development. The antisense nucleic acids can also be used to identify
CC   proteins used in proliferation, to express these proteins, and to obtain
CC   antibodies capable of binding to the expressed proteins. The proteins can
CC   be used to screen compounds in rational drug discovery programmes. The
CC   antisense nucleic acid sequence is also useful to screen for homologous
CC   nucleic acids which are required for cell proliferation in a wide variety
CC   of organisms. The present sequence represents an essential prokaryotic
CC   cellular proliferation protein. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 360 AA;

  Query Match             69.0%;  Score 1268;  DB 4;  Length 360;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-130;
  Matches  239;  Conservative   51;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||

Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|||| || |||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| ||||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:|||||:||||||||||:::|||||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 10
AAG98403
ID   AAG98403 standard; protein; 360 AA.
XX
AC   AAG98403;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Escherichia coli protein sequence SEQ ID NO:451.
XX
KW   Escherichia coli; identification; proliferation; microorganism;
KW   antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW   bacterial growth inhibition.
XX
OS   Escherichia coli.
XX
PN   WO200148209-A2.
XX
PD   05-JUL-2001.
XX
PF   19-DEC-2000; 2000WO-US034419.
XX
PR   23-DEC-1999;   99US-0173005P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen KL,  Zyskind JW;
XX
DR   WPI; 2001-457376/49.
DR   N-PSDB; AAH81459.
XX
PT   Novel nucleic acids encoding proteins required for Escherichia coli
PT   proliferation, useful for screening for antimicrobial agents.
XX
PS   Claim 19; Page 570; 596pp; English.
XX
CC   The present invention describes a purified or isolated nucleic acid
CC   sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC   given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC   microorganism is capable of inhibiting proliferation of a microorganism.
CC   (I) have antibacterial and antibiotic activities, and can be used in gene
CC   therapy. Expression of (I) in a microorganism inhibits proliferation of
CC   the microorganism, and the manufactured antibiotic is useful for reducing
CC   the activity or level of a gene product required for proliferation of a
CC   microorganism in a subject, specifically humans. The nucleic acids that
CC   inhibit bacterial growth or proliferation can be used as antisense
CC   therapeutics for killing bacteria. In addition to therapeutic
CC   applications, the nucleic acid sequences complementary to sequences
CC   required for proliferation can be used as diagnostic tools. For example,
CC   nucleic acid probes complementary to proliferation-required sequences
```

```
CC   that are specific for particular species of microorganisms can be used as
CC   probes to identify particular microorganism species in clinical
CC   specimens. AAH81295 to AAH81487 encode the Escherichia coli proteins
CC   given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent
CC   oligonucleotides, which are used in the exemplification of the present
CC   invention
XX
SQ   Sequence 360 AA;

  Query Match             69.0%;  Score 1268;  DB 4;  Length 360;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-130;
  Matches  239;  Conservative   51;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 11
ABU28488
ID   ABU28488 standard; protein; 360 AA.
XX
AC   ABU28488;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #14015.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Escherichia coli.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA32358.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 56412; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 360 AA;

  Query Match             69.0%;  Score 1268;  DB 6;  Length 360;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-130;
  Matches  239;  Conservative   51;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIAHLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|||| || ||||  || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLVGYGVIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQLVVPFFKDVMPQLGLFYILLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| |||   :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWATGNMNFASYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360


RESULT 12
ABU49996
ID   ABU49996 standard; protein; 360 AA.
XX
AC   ABU49996;
```

XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #35523.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Yersinia pestis.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA53866.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 77920; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 68.9%; Score 1265; DB 6; Length 360;
Best Local Similarity 65.6%; Pred. No. 2.5e-130;
Matches 236; Conservative 54; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: ||||| :|| || || ||| | |:|:|||| :|||:|| :|  ||  |||| ||||
Db          1 MLVWLAEYLVKFYSGFNVFSYLTFRAIVSLLTALFISLWMGPHLIAWLQKLQIGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||| :|| :| || |:||  || ||| || : : :| :|::||||||
Db         61 GPESHFSKRGTPTMGGLMILFSITISVLMWAYPSNPYVWCVLFILIGYGIVGFIDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : ||::|| :|||||||:  : ||  :|   :     | |:||  | :  |||: :|:|
Db        121 VRKNTKGLIARWKYFWQSIIALAAAFTMYSIGKDTSATELVVPFFKDIMPQLGLLYVLLA 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||:||||||||||||||||||| ||    :  : :||| || || || :  ||||:
Db        181 YFVIVGTSNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNVNFAAYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| |||||||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGTIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 13
ABU48094
ID ABU48094 standard; protein; 360 AA.
XX
AC ABU48094;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #33621.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Salmomella typhi.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA51964.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 76018; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding

CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 68.8%; Score 1264; DB 6; Length 360;
Best Local Similarity 66.1%; Pred. No. 3.2e-130;
Matches 238; Conservative 52; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||||| :||||| || ||||   || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|:
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
              ||||||: |||||||||||||||||| ||    :  : :||: || || || :  ||||:
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300
              : | |:|||||||||||||||||||||:|||| ||| |||::||| :| |||||||:|
Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360
              ||||::|| |||| |:|:|||||||||:||||||:||||||||||:::|||||||||:|
Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 14
ABU47262
ID ABU47262 standard; protein; 360 AA.
XX
AC ABU47262;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #32789.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Salmomella typhimurium.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA51132.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 75186; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 360 AA;

Query Match 68.6%; Score 1261; DB 6; Length 360;
Best Local Similarity 65.8%; Pred. No. 6.8e-130;
Matches 237; Conservative 53; Mismatches 70; Indels 0; Gaps 0;

```
Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60
              ||: |||:| ::| || || ||| | |:|:|||| :|||:|| ||  ||    || ||||
Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120
              ||:|| ||:|||||||| :|||:| || ||||   || ||| |||| : :| ||:|||||||
Db         61 GPESHFSKRGTPTMGGIMILTSIVISVLLWAYPSNPYVWCVLVVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180
              : |:::|| :|||||| ||  :| |  ||:  :    | |:||  | |  |||:|:::|:
Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240
```

```
Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360

Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLKVR 360
```

RESULT 15
AAU38482
ID AAU38482 standard; protein; 360 AA.
XX
AC AAU38482;
XX
DT 14-FEB-2002 (first entry)
XX
DE Salmonella typhi cellular proliferation protein #373.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS56341.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 14075; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 360 AA;

```
  Query Match             68.5%;  Score 1258;  DB 4;  Length 360;
  Best Local Similarity   65.8%;  Pred. No. 1.5e-129;
  Matches  237;  Conservative   52;  Mismatches   71;  Indels    0;  Gaps    0;

Qy          1 MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWMIRTLQIPQIGQAVRND 60

Db          1 MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRMIARLQKLSFGQVVRND 60

Qy         61 GPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVLVVTLLFGAIGWVDDYRKV 120

Db         61 GPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVLVVLIGYGIIGFVDDYRKV 120

Qy        121 IEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTLIVPMLKSVEIQLGIFFVVLT 180

Db        121 VRKDTKGLIARWKYFWMSVIALGVAFALYLVGKDTPATQLVVPFFKDVMPQLGLFYILLS 180

Qy        181 YFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYLSGNVKFAEYLLIPNVPGAGELI 240

Db        181 YFVIVGTGNAVNLTDGLDGLAIMPTVFVAAGFALVAWATGNMNFANYLHIPYLRHAGELV 240

Qy        241 VFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAALGTIAVIVRQEIVLFIMGGVFVME 300

Db        241 IVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGALGIIAVLLRQEFLLVIMGGVFVVE 300

Qy        301 TLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWPDPRVIVRFWIITVILVLIGLATLKLR 360

Db        301 TLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWPEPRVIVRFWIISLMLVLIGLATLXVR 360
```

Search completed: May 7, 2004, 08:39:26
Job time : 63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 13:18:04 ; Search time 3219 Seconds
(without alignments)
10046.829 Million cell updates/sec

Title: US-10-089-787-1
Perfect score: 1083
Sequence: 1 atgctcctgctgctggccga.....ccaccttgaagctgcgttga 1083

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	629.4	58.1	1293	28	BZ568926	BZ568926 pacs2-164
	2	580.8	53.6	1253	28	BZ579253	BZ579253 msh2_6268
c	3	564	52.1	754	28	BZ550302	BZ550302 pacs1-60_
	4	324.4	30.0	590	28	CC140622	CC140622 NDL.67G2.
	5	316.4	29.2	636	29	CC857250	CC857250 NDL.101A2
	6	196	18.1	1026	28	BZ559351	BZ559351 pacs2-164
	7	194.8	18.0	430	29	CC868241	CC868241 NDL.44D16
	8	79	7.3	158	10	AW863470	AW863470 MR3-SN000
c	9	72	6.6	891	28	AZ683582	AZ683582 ENTKK47TR
c	10	69.8	6.4	332	28	AZ578373	AZ578373 22b10 Sho
	11	66.8	6.2	816	28	AZ535744	AZ535744 ENTCQ25TR
	12	66.8	6.2	906	28	BH153606	BH153606 ENTTS83TF
	13	65.8	6.1	529	14	CF143195	CF143195 UI-HF-BR0
c	14	64.6	6.0	890	28	BH146886	BH146886 ENTPK48TF
c	15	63	5.8	786	29	CG311108	CG311108 OGYBG71TH
c	16	63	5.8	818	14	CB633135	CB633135 OSIIEb11M
c	17	63	5.8	818	14	CB633177	CB633177 OSIIEb11N
c	18	63	5.8	834	14	CB649819	CB649819 OSJNEb13P
	19	62.6	5.8	406	14	CF142680	CF142680 UI-HF-BR0
	20	61.8	5.7	480	9	AL828696	AL828696 AL828696
	21	61.2	5.7	1300	11	AY106071	AY106071 Zea mays
c	22	61	5.6	597	28	BZ652916	BZ652916 OGAMC66TM
	23	61	5.6	694	14	CD863587	CD863587 AZO1.107E
	24	61	5.6	721	13	BQ620395	BQ620395 TaLr1164E
	25	60.8	5.6	682	12	BJ465499	BJ465499 BJ465499
	26	60.8	5.6	700	9	AL508300	AL508300 AL508300
	27	60.6	5.6	789	29	CG214539	CG214539 OGYBX83TH
c	28	60.6	5.6	887	13	BQ731479	BQ731479 AGENCOURT
	29	60.2	5.6	526	9	AL825814	AL825814 AL825814
	30	60.2	5.6	534	14	CD874750	CD874750 AZO3.102P
	31	60.2	5.6	583	14	CA229130	CA229130 SCAGFL302
	32	60.2	5.6	607	13	BQ620757	BQ620757 TaLr1129F
	33	60.2	5.6	623	13	BQ620376	BQ620376 TaLr1166D
	34	60.2	5.6	664	14	CD875475	CD875475 AZO3.105E
	35	60.2	5.6	675	14	CD875233	CD875233 AZO3.104J
	36	60.2	5.6	687	14	CD874084	CD874084 AZO3.101F
	37	60.2	5.6	695	14	CD863246	CD863246 AZO1.106C
	38	60.2	5.6	698	14	CD863803	CD863803 AZO1.108A
	39	60.2	5.6	699	14	CD874586	CD874586 AZO3.102I
	40	60.2	5.6	718	13	BQ789402	BQ789402 WHE4160_G
	41	59.4	5.5	587	14	CA220741	CA220741 SCBGFL405
	42	59.4	5.5	625	14	CD219591	CD219591 CCC1_57_C
	43	59.4	5.5	658	14	CD222026	CD222026 CCC1_1_F0
	44	59.4	5.5	670	14	CD227954	CD227954 CCC1_56_H
c	45	59.4	5.5	683	14	CF847209	CF847209 psHB042xD

ALIGNMENTS

RESULT 1
BZ568926
LOCUS BZ568926 1293 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_8156.x1 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_8156, genomic survey sequence.
ACCESSION BZ568926
VERSION BZ568926.1 GI:27202739
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1293)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Psedomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES Location/Qualifiers

```
     source          1. .1293
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_8156"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             58.1%;  Score 629.4;  DB 28;  Length 1293;
  Best Local Similarity   97.2%;  Pred. No. 1.2e-86;
  Matches  650;  Conservative    0;  Mismatches   18;  Indels    1;  Gaps    1;

Qy        405 CTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCC 464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         90 CTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCC 149

Qy        465 GATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTT 524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        150 GATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTT 209

Qy        525 CTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGA 584
              |||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||
Db        210 CTTCGTGGTCCTGACCTACTTCGTCATCGTCGGTTCGAGCAATGCAGTGAACCTCACCGA 269

Qy        585 CGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTT 644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        270 CGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTT 329

Qy        645 CTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGG 704
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        330 CTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGTTGATTCCCAACGTACCGGG 389

Qy        705 CGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTG 764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        390 CGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTG 449

Qy        765 GTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGC 824
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        450 GTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGTTGGCGCTGGGCGC 509

Qy        825 CGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGT-G 883
              |||||||||||||||||||||||||||||||| |||||||||||||||||||||| |
Db        510 CGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAAATCGTGCTGTTCATCATGGGTGG 569

Qy        884 GGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCG 943
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        570 GGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTNCTTCAAGCTGACCG 629

Qy        944 GACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGG 1003
              |||||||||||||||| |||||||||||||||||||||||||||||||||||| |||||
Db        630 GACGCCGCGTCTTCCGCATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTTGCCGG 689

Qy       1004 ACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCG 1063
              | |||||||||||||||||||  |||||||||||||||||||||||||||||| |
Db        690 AGCCGCGCGTGATCGTGCGCTTCTTAATCATCACCGTGATCCTGGTGCTGATCGGNCCTC 749

Qy       1064 CCACCTTGA 1072
                |  |||||
Db        750 GCCACTTGA 758


RESULT 2
BZ579253
LOCUS       BZ579253               1253 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  msh2_6268.x1 msh Pseudomonas aeruginosa genomic clone msh2_6268,
            genomic survey sequence.
ACCESSION   BZ579253
VERSION     BZ579253.1  GI:27214314
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 1253)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1253
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="MSH"
                     /db_xref="taxon:287"
                     /clone="msh2_6268"
                     /clone_lib="msh"
                     /note="Environmental isolate. Whole genomic shotgun
                     library."
ORIGIN

  Query Match             53.6%;  Score 580.8;  DB 28;  Length 1253;
  Best Local Similarity   95.9%;  Pred. No. 3.4e-79;
  Matches  636;  Conservative    0;  Mismatches   22;  Indels    5;  Gaps    4;

Qy        260 CGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTAGTGCTGGTCGTTACCCTGC 319
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         97 CCCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTAGTGCTGGTCGTTACCCTGC 156

Qy        320 TGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTGATCGAGAAGAACTCCCGTG 379
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        157 TGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTGATCGAGAAGAACTCCCGTG 216

Qy        380 GCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGT 439
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        217 GCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGT 276

Qy        440 TCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGA 499
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        277 TCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGA 336

Qy        500 GCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCT 559
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||| |
Db        337 GCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACTTACTTCGTCATCGTCGGTT 396

Qy        560 CGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAA 619
              ||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        397 CGAGCAATGCAGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAA 456

Qy        620 TGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGT 679
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 TGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGT 516

Qy        680 ACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGG 739
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        517 ACCTGTTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGG 576

Qy        740 TCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCG 799
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        577 TCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCG 636

Qy        800 ACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGC-GGTGATCGTGCGCCAG 858
              ||||  ||||| |||||||||||||||||||||||||||||| ||||||||||||||||
Db        637 ACGTNCGCGCGTTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGGTGATCGTGCGCCAG 696
```

```
Qy        859 G-AGATCGTGC-TGTTCATCATGGGTGGGG--TGTTCGTCATGGAAACCCTCTCGGTGAT 914
              | | ||||||| || |||| |||||||||     | |||||||||| || |||   | |
Db        697 GAANATCGTGCTTGNTCATTATGGGTGGGNGTGNTCCGTCATGGAAAACCCCTCTTCGGT 756

Qy        915 GAT 917
              |||
Db        757 GAT 759


RESULT 3
BZ550302/c
LOCUS       BZ550302                754 bp    DNA     linear   GSS 17-DEC-2002
DEFINITION  pacs1-60_2592.y2 pacs1-60 Pseudomonas aeruginosa genomic clone
            pacs1-60_2592, genomic survey sequence.
ACCESSION   BZ550302
VERSION     BZ550302.1  GI:27153883
KEYWORDS    GSS.
SOURCE      Pseudomonas aeruginosa
  ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .754
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="1-60"
                     /db_xref="taxon:287"
                     /clone="pacs1-60_2592"
                     /clone_lib="pacs1-60"
                     /note="clinical isolate 1-60 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             52.1%;  Score 564;  DB 28;  Length 754;
  Best Local Similarity   88.5%;  Pred. No. 1.1e-76;
  Matches  634;  Conservative    0;  Mismatches   80;  Indels    2;  Gaps    2;

Qy        205 AAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTTACCGCCATAGCCATCAGCACGCTG 264
              || ||    | ||||||||||||||||  ||||| |||| ||||||||||||||| || ||
Db        753 AAAGGGCATCTGACCATGGGCGGCGCCTAGATCC-TACCCCCATAGCCATCAGCCCGTTG 695

Qy        265 CTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTC 324
              |||||||||||||||| |||||| |||||||||  |||||||||| |||  |||| ||
Db        694 CTGTGGGCGGATCTTTACAACCGTTACGTGTGGTAAGTGCTGGTCGAAACCTAGCTGATC 635

Qy        325 GGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTGATCGAGAAGAACTCCCGTGGCCTG 384
              || ||||  | |||||| ||| |||||| |||||||||||||||||||||| |||||||
Db        634 GGAGCCACGGACTGGGTGGACAACTACCTCAAGGTGATCGAGAAGAACTCCCATGGCCTG 575

Qy        385 CCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTC 444
              || |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        574 CCTAGCCTCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTC 515

Qy        445 TACATGACTGCCGAAACCCCGATCGAGACCACCCT-GATCGTGCCGATGCTGAAGAGCGT 503
              |||||||||||||||| |||||||||||||||| |||||||||||||||||||||||||
Db        514 TACATGACTGCCGAAACTCCGATCGAGACCACCCTAGATCGTGCCGATGCTGAAGAGCGT 455

Qy        504 CGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAG 563
              |||||||||||||||||||||||||||||||||| |||||||||||||||| |||||

Db        454 CGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACTTACTTCGTCATCGTCGGTTCGAG 395

Qy        564 CAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGT 623
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        394 CAATGCAGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGT 335

Qy        624 TGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCT 683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        334 TGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCT 275

Qy        684 GCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGG 743
              | |||||||| |||||||||||||||||||||||||||||| |||||||| ||||||||
Db        274 GTTGATTCCCTACGTACCGGGCGCCGGCGAGCTGATCGTGTACTGCGCCGCACTGGTCGG 215

Qy        744 CGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGT 803
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        214 CGCCGGCCTCGGCTTCCTCTGGTACAACACCTATCCGGCGCAGGTCTTCATGGGCGACGT 155

Qy        804 CGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGAT 863
              ||||||| |||||||||||||||||||||||||||||||||||||| ||||| | ||
Db        154 CGGCGCGTTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGAGCGCCGAGGGGG 95

Qy        864 CGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCC 919
                  |   ||  | | ||    || ||      | |  ||   | ||| |||
Db         94 ATCCTCATAGTTCTAGAGCGGCCGCCTCCTCGGATGGAGCTCCTGCTTTATGTTCC 39


RESULT 4
CC140622
LOCUS       CC140622                590 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.67G2.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.67G2, genomic survey sequence.
ACCESSION   CC140622
VERSION     CC140622.1  GI:30009677
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 590)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .590
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.67G2"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
                     by David Severson at the University of Notre Dame and
                     Hongbin Zhang"
ORIGIN

  Query Match             30.0%;  Score 324.4;  DB 28;  Length 590;
  Best Local Similarity   73.1%;  Pred. No. 4.9e-40;
  Matches  431;  Conservative    0;  Mismatches  156;  Indels    3;  Gaps    1;

Qy        271 GCGGATCTTTCCAACCGCTACGTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCC 330
```

```
Db          1 GCCGACCTGCGCAACCGCTACGTGTGGCTGGTGCTGGCGGTGATGCTGTGCTTCGGCGCC 60

Qy        331 ATCGGCTGGGTAGACGACTACCGCAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGC 390

Db         61 ATCGGCTGGTATGACGACTGGATCAAGATCGTCCGCCGTGACCCGAACGGCCTGAAGTCG 120

Qy        391 CGCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATG 450

Db        121 CGCTGGAAGTACCTGCTGCAGTCGATCTTCGGCCTGGCTGCGGGCCTGTTCCTGTTCTAC 180

Qy        451 ACTGCCGAAACCCCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATC 510

Db        181 ACGGCCGACGTGCCGGCAGCGCTGACCTTCTACATCCCGATGTTCAAGTCGGTCGCGCTG 240

Qy        511 CAGTTG---GGCATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAAT 567

Db        241 CCGCTGGCCGGCATCGGCTTCGTGGCCATCGCCTACTTCTGGATCGTCGGCTTCTCCAAC 300

Qy        568 GCGGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCC 627

Db        301 GCGGTGAACCTGACCGACGGCCTTGACGGCCTGGCGATCATGCCCACCGTGCTGGTGGCC 360

Qy        628 GGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTG 687

Db        361 TGCGCGCTGGGCGTGTTCGCCTACGCCTCGGGCAACGTGGTGTTCGCCAACTACCTGCAG 420

Qy        688 ATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCC 747

Db        421 ATCCCGCAGATCCCGGGCGCCGGCGAGCTGGTCATCATCTGTGCGGCCATCGCCGGCGCG 480

Qy        748 GGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGC 807

Db        481 GGCCTGGGCTTCCTCTGGTTCAACACCTACCCGGCCATGGTCTTCATGGGCGACATCGGC 540

Qy        808 GCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCA 857

Db        541 GCACTGGCACTGGGTGCGGTGCTGGGCACGATCGCGGTGATCACCCGCCA 590
```

RESULT 5
CC857250
LOCUS CC857250 636 bp DNA linear GSS 24-JUL-2003
DEFINITION NDL.101A24.T7 Notre Dame Liverpool Aedes aegypti genomic clone NotreDame Liverpool-101A24, genomic survey sequence.
ACCESSION CC857250
VERSION CC857250.1 GI:33214935
KEYWORDS GSS.
SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes; Stegomyia.
REFERENCE 1 (bases 1 to 636)
AUTHORS Loftus,B., Shetty,J., Knudson,D. and Severson,D.
TITLE BAC end sequencing of Aedes aegypti
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: NDL.101A24.SP6
Contact: Brendan Loftus
Department of Eukaryotic Genomics
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3543
Fax: 301-838-0208
Email: enta@tigr.org
Library was provided by David Severson
Seq primer: T7
Class: BAC ends.
FEATURES Location/Qualifiers
source 1. .636
/organism="Aedes aegypti"
/mol_type="genomic DNA"
/strain="liverpool"
/db_xref="taxon:7159"
/clone="NotreDame Liverpool-101A24"
/clone_lib="Notre Dame Liverpool"
/note="Vector: pECBAC1; Site_1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"
ORIGIN

Query Match 29.2%; Score 316.4; DB 29; Length 636;
Best Local Similarity 72.1%; Pred. No. 8.3e-39;
Matches 454; Conservative 0; Mismatches 171; Indels 5; Gaps 3;

```
Qy        208 GGCACCCCGACCATGGGCGGCGCCCTGATCCTTACCGCCATAGCCATCAGCACGCTGCTG 267

Db          7 GGCACGCCGACCATGGGCGGTTCGCTGATCCTGCTCACCATCACCCTGTCGGTGCTGATG 66

Qy        268 TGGGCGGATCTTTCCAACCGCTACGTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGT 327

Db         67 TGGGCCGACCTGCGCAACCGCTACGTGTGGCTGGTGCTGGCGGTGATGCTGTGCTTCGGC 126

Qy        328 GCCATCGGCTGGGTAGACGACTACCGCAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCG 387

Db        127 GCCATCGGCTGGTATGACGACTGGATCAAGATCGTCCGCCGTGACCCGAACGGCCTGAAG 186

Qy        388 AGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTAC 447

Db        187 TCGCGCTGGAAGTACCTGCTGCAGTCGATCTTCGGCCTGGCTGCGGGCCTGTTCCTGTTC 246

Qy        448 ATGACTGCCGAAACCCCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAG 507

Db        247 TACACGGCCGACGTGCCGGCAGCGCTGACCTTCTACATCCCGATGTTCAAGTCGGTCGCG 306

Qy        508 ATCCAGTTG---GGCATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGC 564

Db        307 CTGCCGCTGGCCGGCATCGGCTTCGTGGCCATCGCCTACTTCTGGATCGTCGGCTTCTCC 366

Qy        565 AATGCGGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTT 624

Db        367 AACGCGGTGAACCTGACCGACGGCCTTGACGGCCTGGCGATCATGCCCACCGTGCTGGTG 426

Qy        625 GCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTG 684

Db        427 GCCTGCGCGCTGGGCGTGTTCGCCTACGCCTCGGGCAACGTGGTGTTCGCCAACTACCTG 486

Qy        685 CTGATTCCCAACGTACCGGGCGCCGGCGAGCT-GATCGTGTTCTGCGCCGCGCTGGTCGG 743

Db        487 CAGATCCCGCAGATCCCGGGCGCCGGCGAGCTGGTTCATCATCTGTGCGGCCATCGCCGG 546

Qy        744 CGC-CGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACG 802

Db        547 CGCGGGGCCTGGGCTTCCTCTGGTTCAACACCTACCCAGCCATGGTCTTCATGGGCGACA 606

Qy        803 TCGGCGCGCTGGCGCTGGGCGCCGCGCTGG 832

Db        607 TCGGCGCACTGGCACTGGGTGCGGTGCTGG 636
```

RESULT 6
BZ559351
LOCUS BZ559351 1026 bp DNA linear GSS 17-DEC-2002
DEFINITION pacs2-164_1435.s1 pacs2-164 Pseudomonas aeruginosa genomic clone pacs2-164_1435, genomic survey sequence.
ACCESSION BZ559351
VERSION BZ559351.1 GI:27175535
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1026)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.

```
  TITLE     Whole-Genome-Sequence variation among multiple isolates of
            Psedomonas aeruginosa library
  JOURNAL   J. Bacteriol. (2002) In press
COMMENT     Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .1026
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_1435"
                     /clone_lib="pacs2-164"
                     /note="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN

  Query Match             18.1%;  Score 196;  DB 28;  Length 1026;
  Best Local Similarity   97.2%;  Pred. No. 2.5e-20;
  Matches  210;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 499

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TACCTGACACTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 559

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        560 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 619

Qy        181 GGTCCGCAGTCGCACCT-GTCGAAGAAGGGCACCCC 215
              ||||||||||||||||| |||||||||||   |||
Db        620 GGTCCGCAGTCGCACCTGGTCGAAGAAGGGGCACCC 655


RESULT 7
CC868241
LOCUS       CC868241                430 bp    DNA     linear   GSS 24-JUL-2003
DEFINITION  NDL.44D16.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
            NotreDame Liverpool-44D16, genomic survey sequence.
ACCESSION   CC868241
VERSION     CC868241.1  GI:33228251
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes;
            Stegomyia.
REFERENCE   1  (bases 1 to 430)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished (2003)
COMMENT     Other_GSSs: NDL.44D16.T7
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .430
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="liverpool"
                     /db_xref="taxon:7159"
                     /clone="NotreDame Liverpool-44D16"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
                     by David Severson at the University of Notre Dame and
                     Hongbin Zhang"
ORIGIN

  Query Match             18.0%;  Score 194.8;  DB 29;  Length 430;
  Best Local Similarity   67.4%;  Pred. No. 3.1e-20;
  Matches  290;  Conservative    0;  Mismatches  137;  Indels    3;  Gaps    1;

Qy        176 ACGACGGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGA 235
              | |||||||||||| | ||  | || |||   |||||||| | ||||||| | ||||
Db          1 AGGACGGTCCGCAGACCCATTTCTCCAAGGCCGGCACCCCTTCAATGGGCGGTTCGCTGA 60

Qy        236 TCCTTACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGT 295
              ||||   | ||||  || |     |||| ||||||| || ||  ||||||||||||||
Db         61 TCCTGCTCACCATCACCCTGTCGGTGCTGATGTGGGCCGACCTGCGCAACCGCTACGTGT 120

Qy        296 GGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCA 355
              || | |||||||  || |  |||  ||||| ||||||||||||  |||||||    ||
Db        121 GGCTGGTGCTGGCGGTGATGCTGTGCTTCGGCGCCATCGGCTGGTATGACGACTGGATCA 180

Qy        356 AGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGG 415
              || |  ||       || |    |||||  |  |||||||||||| |   |||||||
Db        181 AGATCGTCCGCCGTGACCCGAACGGCCTGAAGTCGCGCTGGAAGTACCTGCTGCAGTCGA 240

Qy        416 TGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCA 475
              | |||||| | |  || | | |||||| | |   || ||||    |||  | |   |
Db        241 TCTTCGGCCTGGCTGCGGGCCTGTTCCTGTTCTACACGGCCGACGTGCCGGCAGCGCTGA 300

Qy        476 CCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTG---GGCATCTTCTTCGTGG 532
              || |  | | |||||  | |||  |||| | | | | ||   ||||||  ||||||||
Db        301 CCTTCTACATCCCGATGTTCAAGTCGGTCGCGCTGCCGCTGGCCGGCATCGGCTTCGTGG 360

Qy        533 TCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCG 592
              | |  ||||||||   ||||||||||    | | |||||||||| |||||||| || |
Db        361 CCATCGCCTACTTCTGGATCGTCGGCTTCTCCCACGCGGTGAACCTGACCGACGGCCTTG 420

Qy        593 ACGGCCTGGC 602
              ||||||||||
Db        421 ACGGCCTGGC 430


RESULT 8
AW863470
LOCUS       AW863470                158 bp    mRNA    linear   EST 22-MAY-2000
DEFINITION  MR3-SN0009-100400-206-e02 SN0009 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW863470
VERSION     AW863470.1  GI:7997520
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
```

```
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR3-SN0009-100
            400-206-e02&t3=2000-04-10&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 27
            High quality sequence stop: 158.
FEATURES              Location/Qualifiers
     source           1. .158
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /dev_stage="Adult"
                      /clone_lib="SN0009"
                      /note="Organ: stomach_normal; Vector: puc18; Site_1: SmaI;
                      Site_2: SmaI; A mini-library was made by cloning products
                      derived from  ORESTES PCR (U.S. Letters Patent application
                      No. 196,716 -  Ludwig Institute for Cancer Research)
                      profiles into  the pUC 18 vector. Reverse transcription of
                      tissue mRNA and cDNA amplification were performed under
                      low stringency conditions."
ORIGIN

  Query Match             7.3%;  Score 79;  DB 10;  Length 158;
  Best Local Similarity   77.7%;  Pred. No. 0.013;
  Matches  108;  Conservative    0;  Mismatches   30;  Indels    1;  Gaps    1;

Qy        778 CCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACC 837
              ||  |  |||||||||||||||||| |||||||| |||||  ||  ||||   ||  |||||||||
Db         17 CCTGGCCAGGTCTTCATGGGCGATGTCGGCGCCCTGGCCCTCGGCGGTGCCCTGGGCACC 76

Qy        838 ATCGCGGTGATCGTGCGCCAGGA-GATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCAT 896
              ||||||||  |||   ||  |||||  ||||  |||||   |  |||||||  ||  ||  |||||
Db         77 ATCGCGGTCATCACCCGGCAGGATGATCCTGCTGGCCGTCATGGGCGGCGTCTTCGTTGC 136

Qy        897 GGAAACCCTCTCGGTGATG 915
               |||  |  ||  |||||  |||
Db        137 CGAAGCGCTGTCGGTAATG 155


RESULT 9
AZ683582/c
LOCUS       AZ683582                 891 bp    DNA     linear   GSS 14-DEC-2000
DEFINITION  ENTKK47TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   AZ683582
VERSION     AZ683582.1  GI:11820728
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 891)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 16
            High quality sequence stop: 694.
FEATURES              Location/Qualifiers
     source           1. .891
                      /organism="Entamoeba histolytica"
                      /mol_type="genomic DNA"
                      /strain="HM1:IMSS"
                      /db_xref="taxon:5759"
                      /clone_lib="Entamoeba histolytica Sheared DNA"
                      /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                      Institute for Genomic Research (TIGR), Rockville, MD.
                      Genomic DNA isolated from broth cultures of E. histolytica
                      using a  method described by Clark and Diamond (Clark,
                      C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                      method for isolate identification.  Exp. Parasitol.
                      77:450.). The DNA was mechanically sheared to give a
                      tight size distribution (~2 kb). The v + i method used for
                      the library  construction is described in detail in Smith,
                      H.O. and Venter, J.C.  (Making small insert libraries for
                      whole genome shotgun sequencing  projects. In Genome
                      Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                      Barell, Oxford University Press, 1999)."
ORIGIN

  Query Match             6.6%;  Score 72;  DB 28;  Length 891;
  Best Local Similarity   44.4%;  Pred. No. 0.23;
  Matches  291;  Conservative    0;  Mismatches  365;  Indels    0;  Gaps    0;

Qy        415 GTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACC 474
              ||  ||  |||||  ||  |  |  ||  |||  |||  |  |  |  |  ||  |  |
Db        691 GTCGTCTTCATCGTCGTCTTCATCGTCGTCTTCATCATCTTCTTCATCGTCTTCTTCATC 632

Qy        475 ACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTC 534
              |  |  |||  |     ||  |  |     |  ||   |||   ||   ||||  ||  |||  ||
Db        631 ATCATCATCTTCATCATCTTCATCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC 572

Qy        535 CTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGAC 594
              |  |  |  ||||  |||||  |||  ||     ||     ||    |  |||  |  ||   ||  |
Db        571 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC 512

Qy        595 GGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTG 654
              |  |  |    ||  |     |  ||  |  |  |  ||  |   ||||  ||   ||  |  |
Db        511 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC 452

Qy        655 TCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAG 714
                |  |  |    ||  ||  |  |  |  ||   |  |  |      ||  |  |
Db        451 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC 392

Qy        715 CTGATCGTGTTCTGCGGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACC 774
              |  |||||  ||||     |    |  ||  ||  |  |  ||  |  ||  |||   ||||  |  |
Db        391 ATCATCGTCTTCTAATTCAAATTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCGTC 332

Qy        775 TATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGC 834
              |                |||||||  |  |  |  ||  |     |  |  |  |     |  |  |
Db        331 TTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTC 272

Qy        835 ACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTC 894
              |  ||   ||  |  |    |      |||  |  |||||||      |   |  |||  ||
Db        271 TTCTTCATCGTCTTCTTCATCGTCATCATCATCTTCTTCATCATCATCGTCATCTTCATC 212

Qy        895 ATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTC 954
              ||      |  |||   |  ||  |||     ||     |  ||     |  |  |     |  ||||
Db        211 ATCATCGTCATCTTCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCTTCTTC 152

Qy        955 TTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTG 1014
              ||   ||     |||     |     ||   |  |     ||     |     |  ||
Db        151 ATCATCATCTTCTTCATCATCATCGTCATCTTCATCATCATCTTCTTCATCATCATCTTC 92

Qy       1015 ATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTT 1070
              |||  |   |||||   |||     |  |||  |   |  |||   |  ||   ||  |  |
Db         91 ATCATCGTCTTCTAATTCAAATTCATCATCTTCATCATAATCATCTTCATCATCAT 36
```

```
RESULT 10
AZ578373/c
LOCUS       AZ578373                332 bp    DNA     linear   GSS 08-DEC-2000
DEFINITION  22b10 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
            sp. NGR234 genomic clone 22b10, genomic survey sequence.
ACCESSION   AZ578373
VERSION     AZ578373.1  GI:11605592
KEYWORDS    GSS.
SOURCE      Rhizobium sp. NGR234
  ORGANISM  Rhizobium sp. NGR234
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
REFERENCE   1  (bases 1 to 332)
  AUTHORS   Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
  TITLE     Genetic snapshots of the Rhizobium species NGR234 genome
  JOURNAL   Genome Biol. 1 (6), RESEARCH0014 (2000)
  MEDLINE   21114532
   PUBMED   11178268
COMMENT     Contact: Virginie Viprey
            Laboratoire de Biologie Moleculaire des Plantes Superieures
            University of Geneva
            1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
            Tel: +44(0)1603450000
            Fax: +44(0)1603450045
            Email: virginie.viprey@bbsrc.ac.uk
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .332
                     /organism="Rhizobium sp. NGR234"
                     /mol_type="genomic DNA"
                     /strain="ANU265"
                     /db_xref="taxon:394"
                     /clone="22b10"
                     /clone_lib="Shot-gun genomic library of Rhizobium strain
                     ANU265"
                     /note="Vector: M13; derivative strain of NGR234 cured of
                     pNGR234a"
ORIGIN

  Query Match             6.4%;  Score 69.8;  DB 28;  Length 332;
  Best Local Similarity   71.3%;  Pred. No. 0.39;
  Matches   92;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        955 TTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTG 1014
              ||||  |||||||| ||||| || || |||||   ||| |||||| | ||    |  |||
Db        332 TTCCTGATGGCGCCAATCCACCATCACTTCGAGAAGAAGGGCTGGACCGAGAGCCAGGTG 273

Qy       1015 ATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAG 1074
              |  | ||||||||||||||| | || || |||| | || |||| ||  | ||| | |||
Db        272 GTGATCCGCTTCTGGATCATCGCGGTCATTCTGGCGATGGTCGGTCTTTCGACCCTCAAG 213

Qy       1075 CTGCGTTGA 1083
              ||||| | |
Db        212 CTGCGGTAA 204


RESULT 11
AZ535744
LOCUS       AZ535744                816 bp    DNA     linear   GSS 03-NOV-2000
DEFINITION  ENTCQ25TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   AZ535744
VERSION     AZ535744.1  GI:11092691
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 816)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 36
            High quality sequence stop: 816.
FEATURES             Location/Qualifiers
     source          1. .816
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
ORIGIN

  Query Match             6.2%;  Score 66.8;  DB 28;  Length 816;
  Best Local Similarity   43.0%;  Pred. No. 1.4;
  Matches  329;  Conservative    0;  Mismatches  437;  Indels    0;  Gaps    0;

Qy        291 CGTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTA 350
              |||  |   |  |  |  ||  | | | |  | |||   | | ||  |   ||  |  |
Db         48 CGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCGTCTTCAT 107

Qy        351 CCGCAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCA 410
              |  |   | ||||       |  |  | | |    | |  ||  |    |||||
Db        108 CGTCGTCTTCATCGTCGTCTTCATCGTCGTCTTCATCATCTTCTTCATCGTCTTCTTCAT 167

Qy        411 GTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGA 470
                 |  ||  ||||  |  |  | | ||| ||  | |       |  | |    ||
Db        168 CATCATCATCTTCATCATCTTCATCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTT 227

Qy        471 GACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGT 530
              | || | |  || |  ||  |  | ||  ||  ||  |||| ||||| |
Db        228 CATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTT 287

Qy        531 GGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCT 590
              ||  |   || |||| ||||||||  ||  | ||  |   |  | ||| || |   |
Db        288 CATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTT 347

Qy        591 CGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTA 650
              |  | | |     ||||          |  |  |  || |    |  |||| ||| ||
Db        348 CATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTT 407

Qy        651 CCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGG 710
              | |    | |  | | |   |||  |  ||  |      |  || |||       |  |
Db        408 CATCATCGTCTTCTTCATCATCGTCTTCTAATTCAAATTCATCATCGTCTTCTTCATCAT 467

Qy        711 CGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAA 770
              ||  |  || |  ||| |  |  |  ||  | || | ||  | ||  | ||| ||||
Db        468 CGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCAT 527

Qy        771 CACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCT 830
```

```
Db        528 CGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCATCATCAT 587

Qy        831 GGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTT 890

Db        588 CTTCTTCATCATCATCGTCATCTTCATCATCATCGTCATCTTCATCATCATCTTCTTCAT 647

Qy        891 CGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCG 950

Db        648 CATCATCTTCTTCATCATCATCTTCTTCATCATCATCTTCTTCATCATCATCGTCATCTT 707

Qy        951 CGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCG 1010

Db        708 CATCATCATCTTCTTCATCATCATCTTCATCATCGTCTTCTAATTCAAATTCATCATCTG 767

Qy       1011 CGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATC 1056

Db        768 CATCATCATCATCTTCATCATCATCTTCTTCATCATCGTCTTCATC 813
```

RESULT 12
BH153606

```
LOCUS       BH153606                906 bp    DNA     linear   GSS 24-SEP-2001
DEFINITION  ENTTS83TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   BH153606
VERSION     BH153606.1  GI:15725323
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 906)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 733.
FEATURES             Location/Qualifiers
     source          1. .906
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
ORIGIN
```

```
  Query Match             6.2%;  Score 66.8;  DB 28;  Length 906;
  Best Local Similarity   43.3%;  Pred. No. 1.4;
  Matches  314;  Conservative    0;  Mismatches  412;  Indels    0;  Gaps    0;

Qy        232 CTGATCCTTACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTAC 291

Db         23 CTAATTCAATTCATCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTC 82

Qy        292 GTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTAC 351

Db         83 TTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTC 142

Qy        352 CGCAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAG 411

Db        143 TTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTC 202

Qy        412 TCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAG 471

Db        203 TTCATCGTCTTCTTCATCGTCGTCTTCATCGTCGTCTTCATCGTCGTCTTCATCGTCGTC 262

Qy        472 ACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTG 531

Db        263 TTCATCATCTTCTTCATCGTCTTCTTCATCATCATCATCTTCATCATCTTCATCTTCTTC 322

Qy        532 GTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTC 591

Db        323 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC 382

Qy        592 GACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTAC 651

Db        383 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC 442

Qy        652 CTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGC 711

Db        443 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTC 502

Qy        712 GAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAAC 771

Db        503 ATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTAA 562

Qy        772 ACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTG 831

Db        563 TTCAAATTCATCATCGTCTTCTTCATCATCGTCTTCATCATCGTCCTCTTCATCGTCTTC 622

Qy        832 GGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTC 891

Db        623 TTCATCGTCTTCTTCATCGTCTTCATCATCGTCTTCATCATCGTCTTCTTCATCGTCTTT 682

Qy        892 GTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGC 951

Db        683 CTCATCGGCATCATCATCTTCTTCATCATCATCGTCATCTTCATCATCATCGGCATCCTC 742

Qy        952 GTCTTC 957

Db        743 ATCATC 748
```

RESULT 13
CF143195

```
LOCUS       CF143195                529 bp    mRNA    linear   EST 09-SEP-2003
DEFINITION  UI-HF-BR0p-aqz-b-02-0-UI.r1 NIH_MGC_52 Homo sapiens cDNA clone
            IMAGE:3102243 5', mRNA sequence.
ACCESSION   CF143195
VERSION     CF143195.1  GI:33258639
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 529)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
```

```
   PUBMED   8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Louis Staudt
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/humanfl.html
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1. .529
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3102243"
                     /tissue_type="lymph"
                     /cell_type="germinal center B cells"
                     /cell_line="MGC85"
                     /lab_host="DH10B (LTI)"
                     /clone_lib="NIH_MGC_52"
                     /note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
                     Constructed from size fractionated cytoplasmic mRNA
                     (7.4-9.5kb).  Directionally cloned. Cells provided by
                     Louis M. Staudt, Ph.D.  Library preparation by Maria de
                     Fatima Bonaldo, Ph.D. and  M. Bento Soares, Ph.D. "
ORIGIN

  Query Match             6.1%;  Score 65.8;  DB 14;  Length 529;
  Best Local Similarity   59.3%;  Pred. No. 1.8;
  Matches  112;  Conservative    0;  Mismatches   77;  Indels    0;  Gaps    0;

Qy        670 TTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGC 729
              |||   |  |||  |||| |  ||    | ||||| ||||||  ||   ||  || ||
Db         47 TTCAGTGCCTACGCGCTGCTGGCCTTGCTTGCGGGCGACGGCGAATTGGCTGTCGTCAGC 106

Qy        730 GCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTC 789
              | || ||||  ||| || |  ||||  ||||| | |   |||    |||||||    ||
Db        107 GTCGTGCTGTGCGGTGCGGCACTCGCTTTCCTGTTGCGTAACCTGCATCCGGCCAGCGTG 166

Qy        790 TTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATC 849
              |||||||| ||||  ||  || |  | ||||||  |  |    |  ||||   |||||
Db        167 TTCATGGGGGACGCTGGGTCGATACCACTGGGCTTCCTGGCCGCAGCAATCGGCTTGATC 226

Qy        850 GTGCGCCAG 858
              |  ||| |
Db        227 GGTATCCGG 235


RESULT 14
BH146886/c
LOCUS       BH146886                 890 bp    DNA     linear   GSS 27-AUG-2001
DEFINITION  ENTPK48TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   BH146886
VERSION     BH146886.1  GI:15302963
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 890)
  AUTHORS   Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 17
            High quality sequence stop: 880.
FEATURES             Location/Qualifiers
     source          1. .890
                     /organism="Entamoeba histolytica"
                     /mol_type="genomic DNA"
                     /strain="HM1:IMSS"
                     /db_xref="taxon:5759"
                     /clone_lib="Entamoeba histolytica Sheared DNA"
                     /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
                     Institute for Genomic Research (TIGR), Rockville, MD.
                     Genomic DNA isolated from broth cultures of E. histolytica
                     using a  method described by Clark and Diamond (Clark,
                     C.G., and Diamond, L.S.  (1993) Entamoeba histolytica: a
                     method for isolate identification.  Exp. Parasitol.
                     77:450.). The DNA was mechanically sheared to give a
                     tight size distribution (~2 kb). The v + i method used for
                     the library  construction is described in detail in Smith,
                     H.O. and Venter, J.C.  (Making small insert libraries for
                     whole genome shotgun sequencing  projects. In Genome
                     Sequencing: A Practical Approach, eds. M. Vaudin  and B.
                     Barell, Oxford University Press, 1999)."
ORIGIN

  Query Match             6.0%;  Score 64.6;  DB 28;  Length 890;
  Best Local Similarity   43.1%;  Pred. No. 3.1;
  Matches  313;  Conservative    0;  Mismatches  414;  Indels    0;  Gaps    0;

Qy        231 CCTGATCCTTACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTA 290
              | | ||| |   | ||| | | ||  ||   | | |     |   | | || |  ||
Db        791 CATCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTT 732

Qy        291 CGTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTA 350
              | |   ||  |  |  ||||  | | |  ||     ||||| ||   |   || ||
Db        731 CTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTT 672

Qy        351 CCGCAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCA 410
              |  ||  ||  ||   |  || | |     | | |  | | | |  |||
Db        671 CTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTT 612

Qy        411 GTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGA 470
              |   |  ||  | ||  || || | |  || ||  | | |  | ||    |   |||
Db        611 CTTCATCGTCTTCTTCATCGTCGTCTTCATCGTCGTCTTCATCGTCGTCTTCATCGTCGT 552

Qy        471 GACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGT 530
              ||  |  ||  |    |  |    | | ||   |||    |   | ||| |||||  |
Db        551 CTTCATCATCTTCTTCATCGTCTTCTTCATCATCATCATCTTCATCATCTTCATCTTCTT 492

Qy        531 GGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCT 590
              ||  |  || |||| |||||||  ||  | ||  |   |  | ||| || |    |
Db        491 CATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTT 432

Qy        591 CGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTA 650
              |  | | |    ||||     | | |  || |   |    |||| ||| ||
Db        431 CATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTT 372

Qy        651 CCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGG 710
              | |   | | | | |  ||| |  | | | |  |  |  |  | | |  | | |
Db        371 CATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTT 312

Qy        711 CGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAA 770
              |   |  || | |||  |  ||   | ||  | || | || | ||  | || ||  ||| |
Db        311 CATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCATCGTCTTCTA 252

Qy        771 CACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCT 830
```

```
Db        251 ATTCAAATTCATCATCGTCTTCTTCATCATCGTCTTCTTCATCGTCTTCTTCATCGTCTT 192

Qy        831 GGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTT 890

Db        191 CTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTTCTTCATCGTCTT 132

Qy        891 CGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCG 950

Db        131 CTTCATCGTCATCATCATCTTCTTCATCATCATCGTCATCTTCATCATCATCGTCATCTT 72

Qy        951 CGTCTTC 957

Db         71 CATCATC 65


RESULT 15
CG311108/c
LOCUS       CG311108                786 bp    DNA     linear   GSS 26-AUG-2003
DEFINITION  OGYBG71TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMa0640K21,
            genomic survey sequence.
ACCESSION   CG311108
VERSION     CG311108.1  GI:34228268
KEYWORDS    GSS.
SOURCE      Zea mays
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 786)
  AUTHORS   Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
            Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
            Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
  TITLE     Consortium for Maize Genomics
  JOURNAL   Unpublished (2002)
COMMENT     Other_GSSs: OGYBG71TV
            Contact: Cathy Whitelaw
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-5843
            Fax: 301-838-0208
            Email: whitelaw@tigr.org
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .786
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMBMa0640K21"
                     /clone_lib="ZM_0.7_1.5_KB"
                     /note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
                     methylation filtered genomic DNA library"
ORIGIN

  Query Match             5.8%;  Score 63;  DB 29;  Length 786;
  Best Local Similarity   46.4%;  Pred. No. 5.3;
  Matches  242;  Conservative    0;  Mismatches  277;  Indels    3;  Gaps    1;

Qy        376 CGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCC 435

Db        759 CGTCGGCTGCTCGTGGAGCGGCGGTACCGCCCTCAGCTGGTGATGGCGGTGGCGATCCCT 700

Qy        436 GTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTGATCGTGCCGATGCTG 495

Db        699 TTCTTTCAGCAGGTGACGGGGATCAACGCGATCGCGTTCTACGCGCCGGTGCTGCTGCGC 640

Qy        496 AAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTC 555

Db        639 ACCATCGGCATGGGCGAGAGCGCGTCGCTGCTGTCGGCGGTGGTGACGGGCGTCGTGGGC 580

Qy        556 GGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACG 615

Db        579 GTGGCCTCCACCTCCGCGTCCATGCTCGCCG---TGGACCGCTTTGGGCGCCGCACGCTG 523

Qy        616 GTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCC 675

Db        522 TTCCTGGCCGGCGGCGCGCAGATGCTGGCGTCGCAGGTGCTGATCGGCGCCATCATGGCC 463

Qy        676 GAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCG 735

Db        462 GCGGAGCTGCGGGACAGCGGCGGCGTGGGGAAGGCCTGGGCGGGGGTGCTCATCCTGCTC 403

Qy        736 CTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATG 795

Db        402 ATCGCCGTGTACGTGGCCGGGTTCGGCTGGTCCTGGGGCCCCCTCGGGTGGCTGGTCCCC 343

Qy        796 GGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGC 855

Db        342 AGCGAGATCTTCCCGCTGGAGGTGCGCGCCGCCGGGCAGAGCGTCACGGTGGCGGTGAGC 283

Qy        856 CAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATG 897

Db        282 TTCGCGTTCACCGTGTTCGTGGCGCAGGCGTTCCTGTCCATG 241


Search completed: May  9, 2004, 15:45:41
Job time : 3224 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 13:30:53 ; Search time 524 Seconds
(without alignments)
9362.558 Million cell updates/sec

Title: US-10-089-787-1
Perfect score: 1083
Sequence: 1 atgctcctgctgctggccga.....ccaccttgaagctgcgttga 1083

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2941586 seqs, 2264995651 residues

Total number of hits satisfying chosen parameters: 5883172

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1078.2	99.6	1083	9	US-09-815-242-7913	Sequence 7913, Ap
2	1078.2	99.6	1083	13	US-10-282-122A-30469	Sequence 30469, A
3	819.8	75.7	1080	13	US-10-282-122A-31717	Sequence 31717, A
4	767	70.8	1080	13	US-10-282-122A-33529	Sequence 33529, A
5	540.8	49.9	1080	13	US-10-282-122A-19657	Sequence 19657, A
6	539	49.8	1083	9	US-09-815-242-7594	Sequence 7594, Ap
7	537.6	49.6	1080	13	US-10-282-122A-23402	Sequence 23402, A
8	518.2	47.8	1083	13	US-10-282-122A-39002	Sequence 39002, A
9	518.2	47.8	1083	13	US-10-282-122A-39834	Sequence 39834, A
10	517.2	47.8	1083	9	US-09-815-242-9977	Sequence 9977, Ap
11	502.2	46.4	1083	9	US-09-741-669-258	Sequence 258, App
12	502.2	46.4	1083	9	US-09-815-242-5931	Sequence 5931, Ap
13	502.2	46.4	1083	13	US-10-282-122A-20228	Sequence 20228, A
14	490.6	45.3	1060	13	US-10-282-122A-37049	Sequence 37049, A
15	480.4	44.4	1167	13	US-10-282-122A-15193	Sequence 15193, A
16	470.2	43.4	1083	13	US-10-282-122A-41275	Sequence 41275, A
17	460.6	42.5	1083	13	US-10-282-122A-41736	Sequence 41736, A
18	440.4	40.7	1167	13	US-10-282-122A-13829	Sequence 13829, A
19	412.6	38.1	1167	13	US-10-282-122A-12921	Sequence 12921, A
20	410	37.9	1083	13	US-10-282-122A-29880	Sequence 29880, A
21	409.2	37.8	1080	13	US-10-282-122A-28946	Sequence 28946, A
22	388.6	35.9	1083	13	US-10-282-122A-30712	Sequence 30712, A
23	368	34.0	1080	13	US-10-282-122A-32630	Sequence 32630, A
24	361.4	33.4	1083	9	US-09-815-242-7089	Sequence 7089, Ap
25	361.4	33.4	1083	13	US-10-282-122A-22235	Sequence 22235, A
26	361.4	33.4	1830121	15	US-10-329-960-1	Sequence 1, Appli
27	361.4	33.4	1830121	16	US-10-329-670-1	Sequence 1, Appli
28	326	30.1	645	13	US-10-282-122A-11904	Sequence 11904, A
29	253.4	23.4	1116	13	US-10-282-122A-8532	Sequence 8532, Ap
30	242.4	22.4	717	13	US-10-282-122A-25283	Sequence 25283, A
31	232.2	21.4	100848	13	US-10-672-787-39	Sequence 39, Appl
32	231.4	21.4	1095	13	US-10-282-122A-26874	Sequence 26874, A
33	220.6	20.4	1071	15	US-10-156-761-6096	Sequence 6096, Ap
34	220.6	20.4	9025608	15	US-10-156-761-1	Sequence 1, Appli
35	205.8	19.0	1077	13	US-10-282-122A-25879	Sequence 25879, A
36	175.4	16.2	1080	9	US-09-712-363-80	Sequence 80, Appl
37	175.4	16.2	1080	13	US-10-282-122A-28470	Sequence 28470, A
c 38	174.2	16.1	640681	9	US-09-790-988-1	Sequence 1, Appli
39	173.2	16.0	1062	13	US-10-335-977-836	Sequence 836, App
40	171.8	15.9	1128	13	US-10-335-977-837	Sequence 837, App
41	160.4	14.8	1062	9	US-09-815-242-7231	Sequence 7231, Ap
42	160.4	14.8	1062	13	US-10-282-122A-22534	Sequence 22534, A
43	153	14.1	570	13	US-10-335-977-835	Sequence 835, App
44	148.2	13.7	1098	9	US-09-738-626-2368	Sequence 2368, Ap
c 45	148.2	13.7	3309400	9	US-09-738-626-1	Sequence 1, Appli

ALIGNMENTS

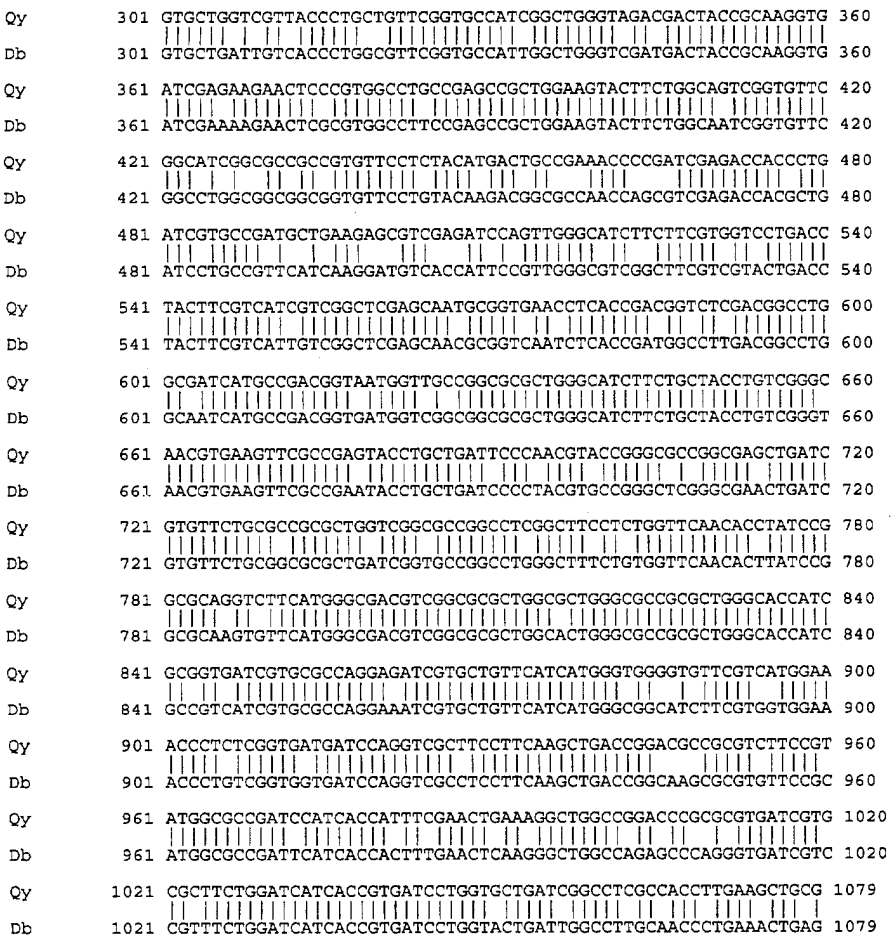
RESULT 1
US-09-815-242-7913
; Sequence 7913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7913
; LENGTH: 1083
; TYPE: DNA

```
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1083)
US-09-815-242-7913

  Query Match             99.6%;  Score 1078.2;  DB 9;  Length 1083;
  Best Local Similarity   99.7%;  Pred. No. 1.6e-270;
  Matches 1080;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATCGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080

Qy       1081 TGA 1083
              |||
Db       1081 TGA 1083


RESULT 2
US-10-282-122A-30469
; Sequence 30469, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30469
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30469

  Query Match             99.6%;  Score 1078.2;  DB 13;  Length 1083;
  Best Local Similarity   99.7%;  Pred. No. 1.6e-270;
  Matches 1080;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
```

```
Db         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTG 360
Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATCGTG 1020
Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Db       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Qy       1081 TGA 1083
Db       1081 TGA 1083
```

RESULT 3

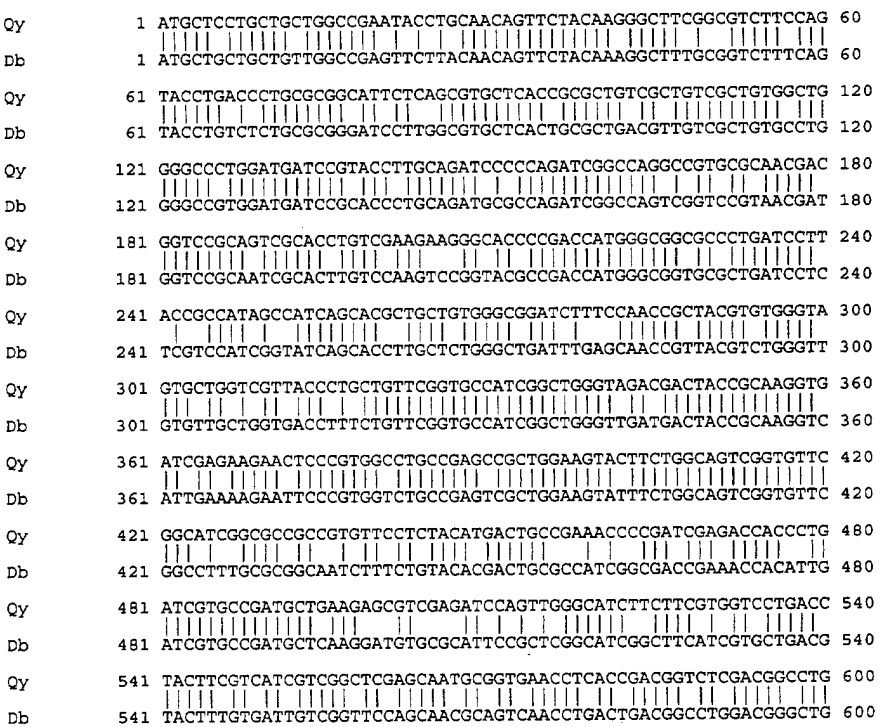
```
US-10-282-122A-31717
; Sequence 31717, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31717
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Pseudomonas putida
US-10-282-122A-31717

  Query Match           75.7%;  Score 819.8;  DB 13;  Length 1080;
  Best Local Similarity  85.0%;  Pred. No. 2.3e-203;
  Matches  917;  Conservative    0;  Mismatches  162;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
Db          1 ATGCTGCTGCTGTTGGCCGAGTATCTGCAACAGTTCCACAAAGGCTTCGCGGTCTTTCAG 60
Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Db         61 TACCTGTCCCTGCGCGGGATTCTTGGTGTACTGACCGCGTTGTCCCTGGCGCTGTGGCTG 120
Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGCCCCTGGATGATCCGTACCCTGCAAATTCGCCAGATCGGCCAGGCCGTGCGTAACGAC 180
Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGCCCGCAATCGCACCTGTCCAAGTCCGGCACCCCGACCATGGGTGGGGCGCTGATCCTG 240
Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 TCGGCCATCGCCGTCAGCACCTTGCTGTGGGCCGACCTGAGCAACCGCTATGTATGGGTT 300
```



```
Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTGCTGATTGTCACCCTGGCGTTCGGTGCCATTGGCTGGGTCGATGACTACCGCAAGGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 ATCGAAAAGAACTCGCGTGGCCTTCCGAGCCGCTGGAAGTACTTCTGGCAATCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GGCCTGGCGGCGGCGGTGTTCCTGTACAAGACGGCGCCAACCAGCGTCGAGACCACGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 ATCCTGCCGTTCATCAAGGATGTCACCATTCCGTTGGGCGTCGGCTTCGTCGTACTGACC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTCGTCATTGTCGGCTCGAGCAACGCGGTCAATCTCACCGATGGCCTTGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCAATCATGCCGACGGTGATGGTCGGCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGT 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db        661 AACGTGAAGTTCGCCGAATACCTGCTGATCCCCTACGTGCCGGGCTCGGGCGAACTGATC 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db        721 GTGTTCTGCGGCGCGCTGATCGGTGCCGGCCTGGGCTTTCTGTGGTTCAACACTTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db        781 GCGCAAGTGTTCATGGGCGACGTCGGCGCGCTGGCACTGGGCGCCGCGCTGGGCACCATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db        841 GCCGTCATCGTGCGCCAGGAAATCGTGCTGTTCATCATGGGCGGCATCTTCGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db        901 ACCCTGTCGGTGGTGATCCAGGTCGCCTCCTTCAAGCTGACCGGCAAGCGCGTGTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db        961 ATGGCGCCGATTCATCACCACTTTGAACTCAAGGGCTGGCCAGAGCCCAGGGTGATCGTC 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCG 1079
Db       1021 CGTTTCTGGATCATCACCGTGATCCTGGTACTGATTGGCCTTGCAACCCTGAAACTGAG 1079
```

RESULT 4
US-10-282-122A-33529
; Sequence 33529, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33529
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-282-122A-33529

Query Match 70.8%; Score 767; DB 13; Length 1080;
Best Local Similarity 81.9%; Pred. No. 1.2e-189;
Matches 884; Conservative 0; Mismatches 195; Indels 0; Gaps 0;



```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
Db          1 ATGCTGCTGCTGTTGGCCGAGTTCTTACAACAGTTCTACAAAGGCTTTGCGGTCTTTCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Db         61 TACCTGTCTCTGCGCGGGATCCTTGGCGTGCTCACTGCGCTGACGTTGTCGCTGTGCCTG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGGCCGTGGATGATCCGCACCCTGCAGATGCGCCAGATCGGCCAGTCGGTCCGTAACGAT 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGTCCGCAATCGCACTTGTCCAAGTCCGGTACGCCGACCATGGGCGGTGCGCTGATCCTC 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 TCGTCCATCGGTATCAGCACCTTGCTCTGGGCTGATTTGAGCAACCGTTACGTCTGGGTT 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTGTTGCTGGTGACCTTTCTGTTCGGTGCCATCGGCTGGGTTGATGACTACCGCAAGGTC 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 ATTGAAAAGAATTCCCGTGGTCTGCCGAGTCGCTGGAAGTATTTCTGGCAGTCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GGCCTTTGCGCGGCAATCTTTCTGTACACGACTGCGCCATCGGCGACCGAAACCACATTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 ATCGTGCCGATGCTCAAGGATGTGCGCATTCCGCTCGGCATCGGCTTCATCGTGCTGACG 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTTGTGATTGTCGGTTCCAGCAACGCAGTCAACCTGACTGACGGCCTGGACGGGCTG 600
```

```
Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCGATCATGCCGACGGTGATGGTCGGCGGTGCGCTGGGCATCTTCTGCTACCTGTCGGGT 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db        661 AACGTGAAATTCGCGGAATACCTGCTGATTCCCTACGTGCCGGGGGCGGGTGAGCTGATT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db        721 GTCTTCTCGGGTGCACTGATTGGTGCGGGGCTCGGGTTTCTCTGGTTCAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db        781 GCTCAGGTGTTCATGGGCGACGTCGGTGCGCTGGCACTGGGCGCTGCGCTGGGCACCATG 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db        841 GCCGTGATCGTGCGTCAGGAAATGGTCCTGTTCATCATGGGCGGCGTGTTCGTGATGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db        901 ACCCTTTCAGTGGTCATTCAGGTTGCGTCTTTCAAACTGACCGGCCGCCGCGTATTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db        961 ATGGCGCCGATTCACCACCACTTTGAACTCAAAGGCTGGCCCGAGCCACGCGTGATCGTC 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCG 1079
Db       1021 CGTTTCTGGATCATCACCGTGATTCTGGTGCTCATCGGCCTTGCCACACTGAAACTGAG 1079
```

RESULT 5
US-10-282-122A-19657
; Sequence 19657, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19657
; LENGTH: 1080
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-10-282-122A-19657

Query Match 49.9%; Score 540.8; DB 13; Length 1080;
Best Local Similarity 68.8%; Pred. No. 7.9e-131;
Matches 743; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCAGGCTTTAACGTCTTTTCG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACTGCGCTGTTCATCTCGTTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGCCCGCGCATGATTGCCCGTCTGCAAAAACTCTCTTTCGGCCAGGTTGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGTCCGGAATCGCACTTCAGCAAGCGCGGCACGCCGACGATGGGCGGGATCATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 ACCGCCATCGTCGTTTCCGTGCTGCTGTGGGCGTATCCCTCCAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTTCTGACGGTACTGGTGGGTTACGGGATCATCGGTTTTGTTGATGATTACCGCAAAGTC 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 GTTCGCAAAGATACTAAGGGCCTGATCGCGCGCTGGAAATATTTCTGGATGTCGGTGATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GCGCTGGGCGTGGCCTTCGCACTGTATCTGGCAGGCAAAGACACGCCAGCCACCGAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 GTCGTGCCGTTCTTTAAGGACGTGATGCCGCAGCTGGGGCTGTTCTACATCCTGCTGGCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTTGTGATTGTTGGTACCGGCAACGCCGTTAACCTGACCGACGGCCTGGATGGTCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCGATTATGCCAACCGTCTTTGTGGCCGCCGGTTTTGCGCTGGTGGCGTGGGCGACCGGT 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db        661 AACATGAACTTTGCAAATTATCTGCACATCCCTTATCTGCGCCACGCGGGTGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db        721 ATCGTCTGTACGGCGATTGTCGGCGCGGGGCTGGGCTTCCTGTGGTTCAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db        781 GCGCAGGTCTTTATGGGGGACGTCGGTTCTCTGGCTTTGGGCGGTGCGCTCGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
```

```
Db        841 GCCGTACTGCTGCGTCAGGAGTTCCTGCTGGTGATCATGGGCGGCGTGTTTGTGGTTGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||| |||||||| | ||||||| |||||||||||||   ||| | ||| ||||||||
Db        901 ACCCTGTCGGTGATTTTGCAGGTCGGTTCCTTCAAGCTGCGCGGTCAGCGCATCTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||| || ||||| | |||||||||||||||||||| ||||||||| |||
Db        961 ATGGCGCCGATTCACCACCACTATGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| || | |||| |||||||||| ||||| || || ||||| | |||
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGGTACGT 1080


RESULT 6
US-09-815-242-7594
; Sequence 7594, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7594
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1083)
US-09-815-242-7594

  Query Match           49.8%;  Score 539;  DB 9;  Length 1083;
  Best Local Similarity   68.6%;  Pred. No. 2.3e-130;
  Matches  743;  Conservative    0;  Mismatches  340;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||| || ||    | |  ||    |||||   ||||||
Db          1 ATGTTAGTATGGCTGGCCGAACACTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCA 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| |||| |||||||||   | ||| ||||| ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCGTTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | ||||||       |||| |  |  |  | |||||| ||| || ||||||
Db        121 GGCCCGCGCATGATCGCCCGTCTGCAAAAACTCGCCTTTGGCCAGGTCGTACGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| |||| || |  ||   || || |||||||||||||   | |||||||
Db        181 GGCCCGGAGTCTCATTTCAGTAAACGCGGTACTCCGACCATGGGCGGGATCATGATCCTC 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||| ||  || |  |  ||||||||||  |||  || ||||  ||||| |||
Db        241 ACCGCGATCACCGTTTCCGTTCTGCTGTGGGCCTATCCATCTAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              || |||  ||     |    | |||  |||||| |  || || || ||||| || ||
Db        301 GTACTGACGGTATTAATCGGCTACGGCATCATCGGTTTCGTTGATGATTACCGTAAAGTC 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |   ||  |  ||   ||||||      |||||||||||| ||||||  ||||||| ||
Db        361 GTGCGCAAAGATACCAAAGGCCTGATCGCCCGCTGGAAGTATTTCTGGATGTCGGTGATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              | | | ||||  |||  |    || ||  || | |  ||     |   | |||  |||
Db        421 GCCCTGGGCGTGGCCTTCGCGCTGTATCTGGCGGGGAAAGATACCCCGGCAACCGAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||||  |  | ||  |||  |  ||| |||| |  ||||   |  | ||| ||
Db        481 GTGGTGCCGTTCTTTAAAGACGTAATGCCGCAGCTGGGGCTGTTCTATATCTTGCTGGCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||||| ||||||||  | |||| || || ||||| |||||||| ||||||||||||
Db        541 TACTTCGTTATCGTCGGTACCGGCAACGCCGTCAACCTGACCGACGGCCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              |||||  |||||||| ||  | || || |  |  |  |  |  |  |  |  | |||
Db        601 GCGATTATGCCGACCGTTTTCGTCGCGGCAGGCTTTGCGCTGGTGGCATGGGCCACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| |||||| | |||||||  || ||  |  | |  ||| |||||||||  |
Db        661 AACATGAATTTCGCCAACTACCTGCATATCCCGTATCTGCGCCACGCTGGCGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  | ||| | ||||| || || || |||||||| ||||||||||||| |||
Db        721 ATCGTCTGTACGGCGATTGTCGGGGCGGGGCTGGGCTTCCTGTGGTTCAACACCTACCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||| ||||| |||||||||||||  | |||||||| |||| |||| ||||||  |||
Db        781 GCGCAAGTCTTTATGGGCGACGTCGGTTCACTGGCGCTCGGCGGCGCGTTGGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| |||||| || ||||| | |||||||| ||||| ||||| |||||
Db        841 GCCGTGCTGCTGCGTCAGGAGTTCCTGCTGGTGATCATGGGCGGGGTTTTCGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||||| ||  | |||||||  ||||| ||||||  ||| |  ||| ||||||||
Db        901 ACGCTGTCGGTTATTCTGCAGGTCGGCTCCTTTAAGCTGCGCGGTCAGCGCATCTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||||||| ||||| |  |||||||||||||||||||| |||||||| ||||||
Db        961 ATGGCGCCGATCCACCACCACTATGAACTGAAAGGCTGGCCGGAGCCGCGCGTTATCGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| ||  |  |||| |||||||||| ||||| || ||  ||||| | |||
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCGACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
              | |
Db       1081 TAA 1083


RESULT 7
US-10-282-122A-23402
; Sequence 23402, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
```

```
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23402
;   LENGTH: 1080
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23402

  Query Match           49.6%;  Score 537.6;  DB 13;  Length 1080;
  Best Local Similarity  68.6%;  Pred. No. 5.4e-130;
  Matches  741;  Conservative    0;  Mismatches  339;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||| || ||    | |  ||    |||||   ||||||
Db          1 ATGTTAGTATGGCTGGCCGAACACTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCA 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| |||| ||||||||||   | ||| ||||| ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCGTTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | ||||||       |||| |  | |    | ||||||| ||| || ||||||
Db        121 GGCCCGCGCATGATCGCCCGTCTGCAAAAACTCGCCTTTGGCCAGGTCGTACGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| |||| ||  |    ||    || || ||||||||||||||   | ||||||||
Db        181 GGCCCGGAGTCTCATTTCAGTAAACGCGGTACTCCGACCATGGGCGGGATCATGATCCTT 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              |||||  ||  ||  |   ||||||||||  |||  ||  ||||  |||||  |||
Db        241 ACCGCGATCACCGTTTCCGTTCTGCTGTGGGCCTATCCATCTAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              || |||   ||      |   | |||   ||||||  |  ||  ||  ||  ||||| || ||
Db        301 GTACTGACGGTATTAATCGGCTACGGCATCATCGGTTTCGTTGATGATTACCGTAAAGTC 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
                |   ||  |  ||   ||||||     ||||||||||||| ||||||  |||||||  ||

Db        361 GTGCGCAAAGATACCAAAGGCCTGATCGCCCGCTGGAAGTATTTCTGGATGTCGGTGATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              | | | ||||  ||| |     || ||  || | |   ||      |     |  |||   |||
Db        421 GCCTTGGGCGTGGCCTTCGCGCTGTATCTGGCGGGGAAAGATACCCCGGCAACCGAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||| |  | ||   |||    |  ||| ||||  | ||||   |  | ||| ||
Db        481 GTGGTGCCGTTCTTTAAAGACGTAATGCCGCAGCTGGGGCTGTTCTATATCTTGCTGGCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||  ||||||||  |  |||| || || ||||| ||||||||  ||||||||||
Db        541 TACTTCGTTATCGTCGGTACCGGCAACGCCGTCAACCTGACCGACGGCCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              |||||  ||||||||  ||  | || ||  |  |  |  |  |  |  |  |  | |||
Db        601 GCGATTATGCCGACCGTTTTCGTCGCGGCAGGCTTTGCGCTGGTGGCATGGGCCACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| ||||||  | |||||  |  || ||   |  |   ||||||||||||| |
Db        661 AACATGAATTTCGCCAACTACCTACATATCCCGTATCTGCGCCACGCCGGCGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  | ||| | ||||| || || || |||||||| |||||||||||||| |||
Db        721 ATCGTCTGTACGGCGATTGTCGGGGCGGGGCTGGGCTTCCTGTGGTTCAACACCTACCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||| ||||| |||||||||||||  | |||||||| |||| |||| ||||||  |||
Db        781 GCGCAAGTCTTTATGGGCGACGTCGGTTCACTGGCGCTCGGCGGCGCGTTGGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| ||||||  || ||||| | |||||||| ||||| |||||  |||||
Db        841 GCCGTGCTGCTGCGTCAGGAGTTCCTGCTGGTGATCATGGGCGGGGTTTTCGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||||| ||  | |||||||  ||||| ||||||  ||| |  ||  |||||||
Db        901 ACGCTGTCGGTTATTCTGCAGGTCGGCTCCTTTAAGCTGCGCGGTCAGCGTATCTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||||||| ||||| |  |||||||||||||||||||| |||||||| ||||||
Db        961 ATGGCGCCGATCCACCACCACTATGAACTGAAAGGCTGGCCGGAGCCGCGCGTTATCGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| ||  |  |||| |||||||||| ||||| || ||  ||||| | |||
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCGACGCTGAAGGTACGT 1080

RESULT 8
US-10-282-122A-39002
; Sequence 39002, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
```

```
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39002
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Salmonella typhimurium
US-10-282-122A-39002

  Query Match             47.8%;  Score 518.2;  DB 13;  Length 1083;
  Best Local Similarity   67.4%;  Pred. No. 5.9e-125;
  Matches  730;  Conservative    0;  Mismatches  353;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  | |||||||| |  ||   | |  ||    |||||  ||||||
Db          1 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| || | ||||||||||   | ||  | ||| ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | ||||||  |    |||| |  | |    | |||||||  || || |||||
Db        121 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| | |||||| |    ||    || || ||||| |||||||||   | |||||||
Db        181 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||  | || |  ||    |   ||| | |||||  | |  || ||||  |||||  |||
Db        241 ACGTCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||||||||  ||    |  | |||    ||||| |  || || ||||||||||| |||
Db        301 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
                |  |||  |  ||   || |||      |||||||| || ||||||  ||||||  ||
Db        361 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  ||||||  ||| |   || ||  |    |  ||  | ||  || ||||   |||
Db        421 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||||  |  ||    ||    |  || |||||  | || | |  |  ||| ||
Db        481 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||| ||||||||||  || | || || |||||||| |||||||| || || || |||
Db        541 TACTTTGTCATCGTCGGTACGGGTAACGCCGTGAACCTGACCGACGGGCTTGATGGTCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||| |||||||| ||  | |||||| ||   | |  |  |   ||    | | ||
Db        601 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCCTGGGCGACCGGG 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| |||||| | |||||||  |||||  |  |||     || |||||||| |
Db        661 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  | ||| | |||||||| ||  | || ||| | ||||| ||||||||||||
Db        721 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||| || |||||||| |||||  |||||||| ||||||  ||| ||||||  |||
Db        781 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGTGCGTTGGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| |||||  | ||||| | |||||||| || || || || |||||
Db        841 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || |||||  ||  | |||||  | ||||||  || ||    |||||  ||  | ||||||
Db        901 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||| |||||||||| | ||||||||||||||||||| |||||||||| |||
Db        961 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| ||  |  |||| |||||||||| ||||| || ||  |||||  | |||
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
              |||
Db       1081 TAA 1083


RESULT 9
US-10-282-122A-39834
; Sequence 39834, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
```

```
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39834
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Salmonella typhi
US-10-282-122A-39834

  Query Match             47.8%;  Score 518.2;  DB 13;  Length 1083;
  Best Local Similarity   67.4%;  Pred. No. 5.9e-125;
  Matches  730;  Conservative    0;  Mismatches  353;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||  |  ||    | |  ||    |||||   ||||||
Db          1 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||  |||||  | |||| |||  ||||| || | ||||||||||    | ||  | ||| ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||  ||  | ||||||  |    ||||  | | |    | |||||||  || || |||||
Db        121 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| | ||||||  |    ||    || || ||||| |||||||||  | |||||||
Db        181 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || || |  ||    |  ||| | ||||  |  |  ||  |||| |||||  |||
Db        241 ACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||| ||      |  | |||    |||||  |  ||  || |||||||||| |||
Db        301 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |   |||  |  ||   || |||      ||||||| || ||||||  ||||||  ||
Db        361 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  ||||||  ||| |  || ||  |   |  ||  | ||  || ||||   |||
Db        421 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||| | | ||   ||   |  || ||||  | || | | |  | ||| ||
Db        481 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||| ||||||||||  || | || || || ||||| ||||||| || || || |||
Db        541 TACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACCGACGGGCTTGATGGTCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||| |||||||| ||  | |||||| ||   | |   |  |     | | ||
Db        601 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCTTGGGCGACCGGG 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| |||||| | |||||| |||||  |  |||   || |||||||| |
Db        661 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  | ||| | |||||||| ||  | || ||| | ||||| ||||||||||
Db        721 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||| || |||||||| |||||  |||||||| |||||| |||| ||||||   |||
Db        781 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGCGCGTTGGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| |||||  |  ||||| | |||||||| || || || ||  |||||
Db        841 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 900
```

```
Qy         901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
               || || |||||  ||  | |||||  | ||||||  || ||   |||||   ||  | ||||||
Db         901 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 960

Qy         961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
               |||||||| |||||||||| |  |||||||||||||||||||| |||||||||| |||
Db         961 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 1020

Qy        1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
               |||||||||| ||  |  |||| |||||||||| ||||| || ||  ||||| | |||
Db        1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGGTACGT 1080

Qy        1081 TGA 1083
               | |
Db        1081 TAA 1083
```

```
RESULT 10
US-09-815-242-9977
; Sequence 9977, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9977
;   LENGTH: 1083
;   TYPE: DNA
;   ORGANISM: Salmonella typhi
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1083)
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1083)
;   OTHER INFORMATION: n = A,T,C or G
US-09-815-242-9977

  Query Match             47.8%;  Score 517.2;  DB 9;  Length 1083;
  Best Local Similarity   67.3%;  Pred. No. 1.1e-124;
  Matches  729;  Conservative    0;  Mismatches  354;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||  |  ||    | |  ||    |||||   ||||||
Db          1 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||  |||||  | |||| |||  |||||  || | ||||||||||    | ||  | ||| ||
```

```
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 120
Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 180
Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 240
Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 ACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 300
Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 360
Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 420
Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 480
Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 540
Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACCGACGGGCTTGATGGTCTG 600
Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCTTGGGCGACCGGG 660
Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db        661 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 720
Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db        721 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 780
Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db        781 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGCGCGTTGGGCATTATC 840
Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db        841 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 900
Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db        901 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 960
Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db        961 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 1020
Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGTACGT 1080
Qy       1081 TGA 1083
Db       1081 TAA 1083

RESULT 11
US-09-741-669-258

; Sequence 258, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
;  APPLICANT:  Forsyth, R. Allyn
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  TITLE OF INVENTION:  Genes identified as required for
;  TITLE OF INVENTION:  proliferation of E. coli
;  FILE REFERENCE:  ELITRA.009A
;  CURRENT APPLICATION NUMBER: US/09/741,669
;  CURRENT FILING DATE:  2000-12-19
;  PRIOR APPLICATION NUMBER:  US 60/173005
;  PRIOR FILING DATE:  1999-12-23
;  NUMBER OF SEQ ID NOS:  481
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 258
;  LENGTH: 1083
;  TYPE: DNA
;  ORGANISM: Escherichia coli
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(1083)
US-09-741-669-258

  Query Match           46.4%;  Score 502.2;  DB 9;  Length 1083;
  Best Local Similarity 66.5%;  Pred. No. 8.6e-121;
  Matches  720;  Conservative    0;  Mismatches  363;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 60
Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 120
Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 180
Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 240
Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 300
Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 360
Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 420
Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 480
Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 540
Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 600
Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 660
Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
```

```
Db        661 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780

Db        721 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840

Db        781 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900

Db        841 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960

Db        901 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020

Db        961 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080

Db       1021 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083

Db       1081 TAA 1083
```

RESULT 12
US-09-815-242-5931
; Sequence 5931, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5931
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1083)

US-09-815-242-5931

Query Match 46.4%; Score 502.2; DB 9; Length 1083;
Best Local Similarity 66.5%; Pred. No. 8.6e-121;
Matches 720; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60

Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120

Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180

Db        121 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240

Db        181 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300

Db        241 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360

Db        301 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420

Db        361 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480

Db        421 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540

Db        481 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600

Db        541 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660

Db        601 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720

Db        661 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780

Db        721 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840

Db        781 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900

Db        841 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960

Db        901 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020

Db        961 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1020
```

```
Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Db       1021 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
Db       1081 TAA 1083
```

RESULT 13
US-10-282-122A-20228
; Sequence 20228, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20228
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-282-122A-20228

Query Match 46.4%; Score 502.2; DB 13; Length 1083;
Best Local Similarity 66.5%; Pred. No. 8.6e-121;
Matches 720; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db        661 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db        721 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db        781 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db        841 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db        901 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db        961 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Db       1021 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
Db       1081 TAA 1083
```

RESULT 14
US-10-282-122A-37049
; Sequence 37049, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37049
; LENGTH: 1060
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-37049

Query Match 45.3%; Score 490.6; DB 13; Length 1060;
Best Local Similarity 68.2%; Pred. No. 8.9e-118;
Matches 710; Conservative 0; Mismatches 329; Indels 2; Gaps 2;

```
Qy         43 GGCTTCGGCGTCTTCCAGTACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTG 102
Db         19 GGCTTTAACGTCTTTTCTTATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTG 78

Qy        103 TCGCTGTCGCTGTGGCTGGGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGC 162
Db         79 TTCATCTCTTTATGGATGGGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGC 138

Qy        163 CAGGCCGTGCGCAACGACGGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATG 222
Db        139 CAGGTGGTACGTAACGATGGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATG 198

Qy        223 GGCGGCGCCCTGATCCTTACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCC 282
Db        199 GGCGGCATCATGATCCTGACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCT 258

Qy        283 AACCGCTACGTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTA 342
Db        259 AACCCGTACGTCTGGTGCGTGCTGGTGGTATTGATTGGCTACGGCATTATCGGTTTTGTC 318

Qy        343 GACGACTACCGCAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTAC 402
Db        319 GATGACTACCACAAAGTGGTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATAT 378

Qy        403 TTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACC 462
Db        379 TTCTGGATGTCGGTTATCGCGCTCGG-GTGGCCTTTGCGCTTTATCTCGTCGGGAAAGAC 437

Qy        463 CCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATC 522
Db        438 ACGCCCGCGACCCAACTGGTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTG 497

Qy        523 TTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACC 582
Db        498 TTTTACATTCTGCTGTCCTACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACC 557

Qy        583 GACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATC 642
Db        558 GACGGGCTTGATGGTCTGGCGATTATGCCGACTGTTTTCGTTGCCGCCGGCTTTGCGCTG 617

Qy        643 TTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCG 702
Db        618 GTGGCCTGGGCGACCGGGAACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGC 677

Qy        703 GGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTC 762
Db        678 TATGCGGGCGAGCTGGTGATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTG 737

Qy        763 TGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGC 822
Db        738 TGGTTTAACACCTATCCGGCGCAGGTTTTTATGGGCGACGTCGGATCGCTGGCGTTGGGC 797

Qy        823 GCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGT 882
Db        798 GGCGCGTTGGGCATTATCGCCGTGCTGCTGCGTCAGGAGTTCCTGCTGGTGATCATGGGC 857

Qy        883 GGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACC 942
Db        858 GGCGTCTTTGTGGTGGAAACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGC 917

Qy        943 GGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCG 1002
Db        918 GGACAGC-TATTTTCCGTATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCG 976

Qy       1003 GACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTC 1062
Db        977 GAACCGCGCGTGATTGTGCGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTG 1036

Qy       1063 GCCACCTTGAAGCTGCGTTGA 1083
Db       1037 GCAACGCTGAAGGTACGTTAA 1057
```

RESULT 15
US-10-282-122A-15193
; Sequence 15193, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

```
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15193
;   LENGTH: 1167
;   TYPE: DNA
;   ORGANISM: Bordetella pertussis
US-10-282-122A-15193

  Query Match             44.4%;  Score 480.4;  DB 13;  Length 1167;
  Best Local Similarity   66.7%;  Pred. No. 4e-115;
  Matches  778;  Conservative    0;  Mismatches  301;  Indels   87;  Gaps    3;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||| ||   | | |||   | |||   |  |     | | | |||| ||| ||
Db          1 ATGCTGCTTGAGATCGCCCGGTGGCTGTCCGACGACGTGCGCGCCATTGGCGTATTCGAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||| | || ||||||| | | ||          || ||||||  | |    |||  | |
Db         61 TACATCACGCTGCGCGCCGTGCTGGCGTGCGCGACGGCGCTGCTGATCGGCCTGGTGGCG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              |||||  |  |||||||   | ||           ||||||||||||||||||||  | ||
Db        121 GGGCCGCGCGTGATCCGCCGCCTGACCGAAATGAAGATCGGCCAGGCCGTGCGCGCCTAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| ||||||||||   |||| |||||| ||||||||||||||||| || |||||
Db        181 GGCCCGGAGTCGCACCTGGTCAAGACGGGCACGCCGACCATGGGCGGCGCGCTCATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              | ||| || |||||||||||||||||||||||||     |||| |||| ||| |||||
Db        241 ATCGCGATCGCCATCAGCACGCTGCTGTGGGCGGACTGGACCAATCGCTTCGTCTGGGTG 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||| || | || ||  |    || ||    ||||| ||| | |||||||| ||||||||
Db        301 GTGTTGCTGGTGACGTTCGGCTTTGGCTGGATCGGGTGGATGGACGACTATCGCAAGGTC 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
               || |      |  ||   ||| |||||  |||   ||||| ||||||||||  ||    ||
Db        361 GTCTACCGCGATCCCGAAGGCATGCCGGCCCGGCAGAAGTTCTTCTGGCAGGCGACCATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACC------------------ 462
              ||  | | ||||||||||||  ||   | | | | |    |
Db        421 GGGCTGGTCGCCGCCGTGTATCTGGCCTTCGCCGTCTCGGCACCCGCCAATACCGAACTG 480

Qy        463 ------------------------------------------------CCGATCGAGACC 474
                                                             ||||  | ||
Db        481 TGGCCGCTCTTCAAGGCATGGGTGGGTAGCGGGTTCACGATGCCGCTGCCGACGCGGGCC 540

Qy        475 ACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTT---CTTCGTG 531
               ||||||||||||  |  | |||   |||      | | ||||| |   |  |||||||
Db        541 GACCTGATCGTGCCGTTCTTCAAGTCGGTCAGCTATCCGCTGGGCGTGCTGGGCTTCGTG 600

Qy        532 GTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTC 591
              | ||||||||    |||||||||| ||| | ||||| || || ||||| ||||| || ||
Db        601 GCCCTGACCTGGGCCGTCATCGTTGGCACCAGCAACGCCGTCAACCTGACCGATGGCCTG 660

Qy        592 GACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTAC 651
              |||||||||| |||||||| |||||  |||||  | |  |||| ||||||||||||   ||||
Db        661 GACGGCCTGGCCATCATGCCTACGGTCATGGTCGGCAGCGCCCTGGGCATCTTCGCCTAC 720

Qy        652 CTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGC 711
               ||   |||  |||   | |   || |||||||||| | ||  ||  | ||  |||| | |
Db        721 GTGGTTGGCCGCGTCGATTATTCCAAGTACCTGCTGTTCCCGTACATTCCCGGCGCGGCC 780

Qy        712 GAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAAC 771
              |||||||| ||| | ||||||||  | |  |||| ||||| | |||||| |||||||||
Db        781 GAGCTGATGGTGCTGTGCGCCGCCATCGGGGGCGCGGGCCTGGCCTTCCTGTGGTTCAAC 840

Qy        772 ACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTG 831
              |||||||||||||  || |||||||||| |||||||||||||||||||||| |||||||
Db        841 GCCTATCCGGCGCAAGTGTTCATGGGCGATGTCGGCGCGCTGGCGCTGGGCGGCGCGCTG 900

Qy        832 GGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTC 891
              |||||||||||||||||||||||||||||| |||||||| |||||||||||||| ||||||
Db        901 GGCACCATCGCGGTGATCGTGCGCCAGGAAATCGTGCTCTTCATCATGGGTGGCGTGTTC 960

Qy        892 GTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACC--------- 942
              ||  | |||||||| |||||  ||| | || ||| | |  |||||     |||
Db        961 GTGGTCGAAACCCTTTCGGTCATGGTGCAAGTCACCTGGTTCAAATACACCAAGCGCAAG 1020

Qy        943 ---------GGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAA 993
                        || || ||  | ||||| |||||||| | ||||||||||||||  ||
Db       1021 TACGGCCAGGGCCGGCGGATATTCCGCATGGCGCCGCTGCATCACCATTTCGAGGTGGGC 1080

Qy        994 GGCTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTG 1053
              ||||||  |||  | |  ||  |||||||| |||||||||||||||| |||| |||||||||
Db       1081 GGCTGGAAGGAAACCCAGGTTGTCGTGCGGTTCTGGATCATCACCATGATGCTGGTGCTG 1140

Qy       1054 ATCGGCCTCGCCACCTTGAAGCTGCG 1079
              |||||||  ||||  |||||  ||||
Db       1141 GTCGGCCTTTCCACACTGAAGTTGCG 1166


Search completed: May  9, 2004, 15:56:26
Job time : 531 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 13:20:48 ; Search time 97 Seconds
(without alignments)
6195.998 Million cell updates/sec

Title: US-10-089-787-1
Perfect score: 1083
Sequence: 1 atgctcctgctgctggccga.....ccaccttgaagctgcgttga 1083

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	1068.6	98.7	1170	4	US-09-252-991A-7929	Sequence 7929, Ap
	2	1068.6	98.7	1404	4	US-09-252-991A-7623	Sequence 7623, Ap
	3	455	42.0	1032	4	US-09-489-039A-5268	Sequence 5268, Ap
	4	366.2	33.8	1083	4	US-09-543-681A-2879	Sequence 2879, Ap
	5	361.4	33.4	1830121	4	US-09-557-884-1	Sequence 1, Appli
	6	361.4	33.4	1830121	4	US-09-643-990A-1	Sequence 1, Appli
	7	250.6	23.1	1119	4	US-09-328-352-2043	Sequence 2043, Ap
	8	233.8	21.6	1107	4	US-09-540-236-1218	Sequence 1218, Ap
	9	232.2	21.4	100848	4	US-09-596-002-39	Sequence 39, Appl
c	10	175.4	16.2	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c	11	175.4	16.2	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c	12	174.2	16.1	640681	4	US-09-790-988-1	Sequence 1, Appli
	13	134	12.4	210	4	US-09-489-039A-5281	Sequence 5281, Ap
	14	133.4	12.3	6911	1	US-08-311-174-4	Sequence 4, Appli
	15	115.2	10.6	2402	4	US-09-221-017B-881	Sequence 881, App
	16	91.2	8.4	1230025	4	US-09-198-452A-1	Sequence 1, Appli
	17	88.4	8.2	996	4	US-09-107-532A-1622	Sequence 1622, Ap
	18	87.4	8.1	981	3	US-08-986-768-1	Sequence 1, Appli
	19	87.4	8.1	981	3	US-08-986-768-3	Sequence 3, Appli
	20	87.4	8.1	3190	3	US-08-986-768-4	Sequence 4, Appli
	21	86	7.9	8148	4	US-08-961-527-11	Sequence 11, Appl
c	22	79.4	7.3	1600	4	US-08-956-171E-159	Sequence 159, App
	23	79.4	7.3	2423	3	US-08-714-918-86	Sequence 86, Appl
	24	79.4	7.3	2423	3	US-09-265-315-86	Sequence 86, Appl
	25	79.4	7.3	2423	3	US-09-265-315-86	Sequence 86, Appl
	26	79.4	7.3	2423	3	US-09-266-417-86	Sequence 86, Appl
	27	79.4	7.3	2423	4	US-09-528-709-86	Sequence 86, Appl
	28	79.4	7.3	2423	4	US-09-527-745-86	Sequence 86, Appl
	29	73.2	6.8	1005	4	US-09-134-001C-1014	Sequence 1014, Ap
	30	72.2	6.7	984	4	US-09-134-000C-1280	Sequence 1280, Ap
	31	71.4	6.6	1416	4	US-09-489-039A-5247	Sequence 5247, Ap
c	32	70.4	6.5	390	3	US-09-197-649-7	Sequence 7, Appli
	33	69.2	6.4	1926	4	US-09-249-585A-4	Sequence 4, Appli
	34	69.2	6.4	1931	2	US-09-130-114-2	Sequence 2, Appli
	35	68.6	6.3	1338	4	US-09-252-991A-16571	Sequence 16571, A
c	36	68.6	6.3	1404	4	US-09-252-991A-16031	Sequence 16031, A
	37	68.6	6.3	2157	4	US-09-252-991A-16462	Sequence 16462, A
	38	68	6.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
	39	68	6.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
	40	64.4	5.9	1392	4	US-09-252-991A-8917	Sequence 8917, Ap
	41	64.4	5.9	1617	4	US-09-252-991A-8802	Sequence 8802, Ap
c	42	64.4	5.9	1956	4	US-09-252-991A-9165	Sequence 9165, Ap
	43	63	5.8	1995	4	US-09-252-991A-14545	Sequence 14545, A
	44	63	5.8	3396	4	US-09-252-991A-14676	Sequence 14676, A
c	45	63	5.8	3444	4	US-09-252-991A-15078	Sequence 15078, A

ALIGNMENTS

RESULT 1
US-09-252-991A-7929/c
; Sequence 7929, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7929
; LENGTH: 1170
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-7929

Query Match 98.7%; Score 1068.6; DB 4; Length 1170;
Best Local Similarity 99.2%; Pred. No. 2.1e-232;
Matches 1074; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1167 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 1108

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1107 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 1048

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1047 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 988

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        987 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 928

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        927 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 868

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        867 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTG 808
```

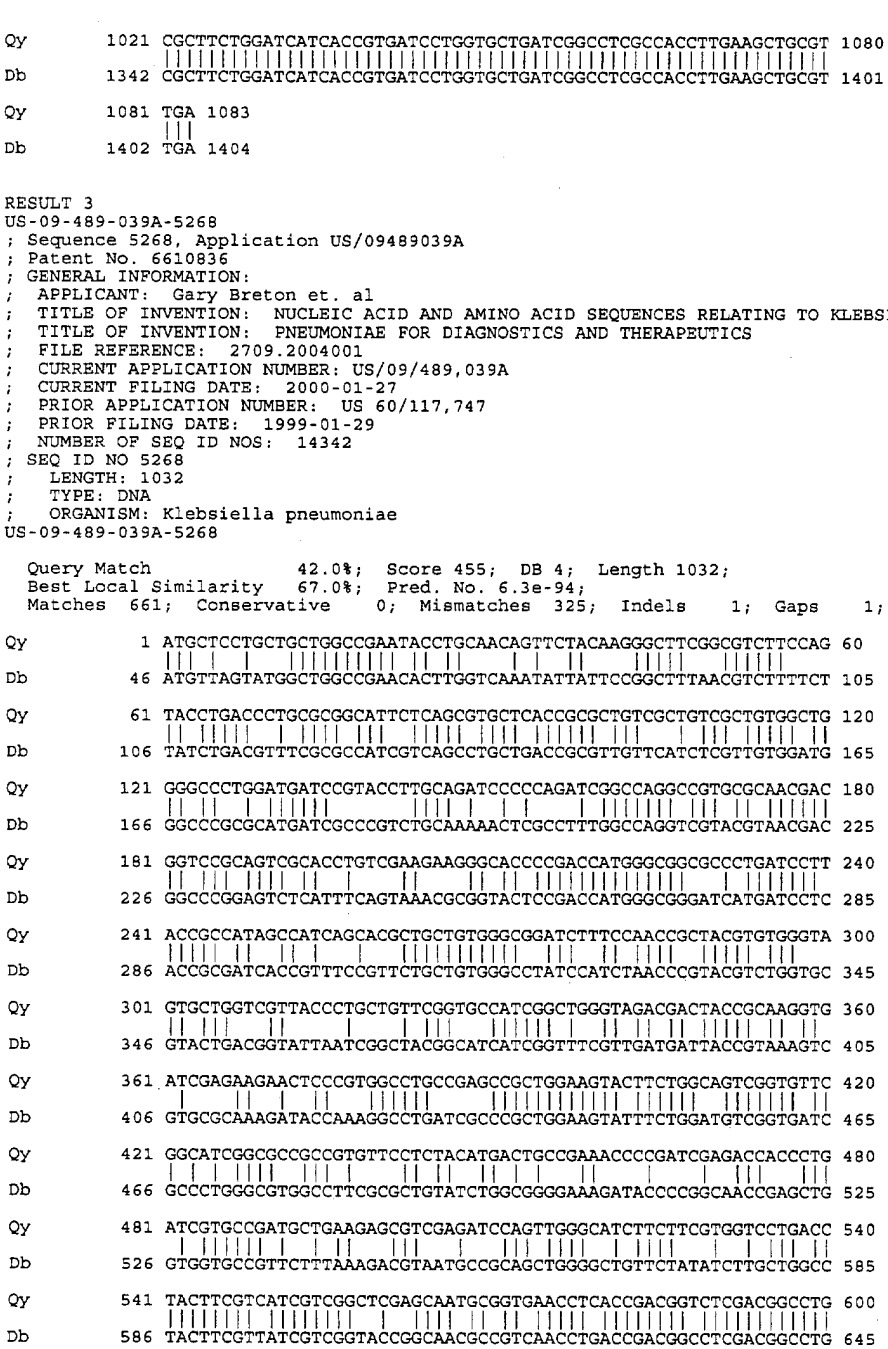
```
Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        807 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 748

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        747 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 688

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        687 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 628

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        627 TACTTCGTCATCGTCGGTTCGAGCAATGCAGTGAACCTCACCGACGGTCTCGACGGCCTG 568

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        567 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 508

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        507 AACGTGAAGTTCGCCGAGTACCTGTTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 448

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 388

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        387 GCGCAGGTCTTCATGGGCGACGTCGGCGCGTTGGCGCTGGGCGCCGCGCTGGGCACCATC 328

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        327 GCGGTGATCGTGCGCCAGGAAATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 268

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGC 208

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        207 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATCGTG 148

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        147 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 88

Qy       1081 TGA 1083
              |||
Db          87 TGA 85
```

RESULT 2
US-09-252-991A-7623
; Sequence 7623, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 7623
; LENGTH: 1404
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-7623

Query Match 98.7%; Score 1068.6; DB 4; Length 1404;
Best Local Similarity 99.2%; Pred. No. 2.2e-232;
Matches 1074; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 381

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 441

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        442 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 501

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        502 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 561

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        562 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 621

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        622 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTG 681

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        682 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 741

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        742 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 801

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        802 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 861

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||||||||||||| |||||||||| ||||||||||||||||||||||||||||||||
Db        862 TACTTCGTCATCGTCGGTTCGAGCAATGCAGTGAACCTCACCGACGGTCTCGACGGCCTG 921

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        922 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 981

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        982 AACGTGAAGTTCGCCGAGTACCTGTTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 1041

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1042 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 1101

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1102 GCGCAGGTCTTCATGGGCGACGTCGGCGCGTTGGCGCTGGGCGCCGCGCTGGGCACCATC 1161

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1162 GCGGTGATCGTGCGCCAGGAAATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 1221

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1222 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGC 1281

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       1282 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATCGTG 1341
```

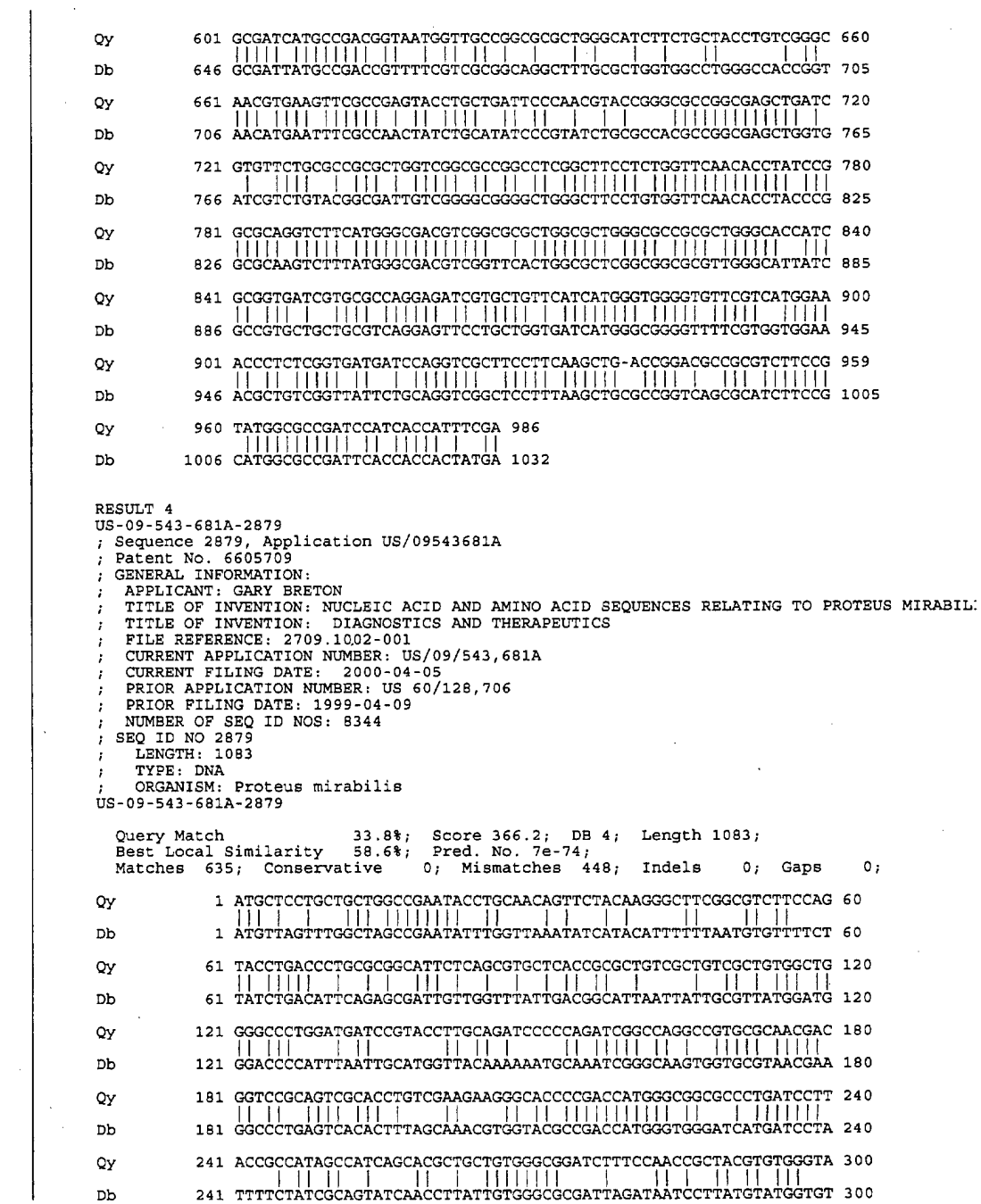


RESULT 3
US-09-489-039A-5268
; Sequence 5268, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 5268
; LENGTH: 1032
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5268

Query Match 42.0%; Score 455; DB 4; Length 1032;
Best Local Similarity 67.0%; Pred. No. 6.3e-94;
Matches 661; Conservative 0; Mismatches 325; Indels 1; Gaps 1;

RESULT 4
US-09-543-681A-2879
; Sequence 2879, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2879
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2879

Query Match 33.8%; Score 366.2; DB 4; Length 1083;
Best Local Similarity 58.6%; Pred. No. 7e-74;
Matches 635; Conservative 0; Mismatches 448; Indels 0; Gaps 0;



```
Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              || ||  | ||      |    | ||   |||||| |   | || || || ||||| ||
Db        301 GTACTCTTGGTGCTAATTGGTTATGGCATCATCGGTTTTATCGATGATTATCGCAAAGTA 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |    ||  |  | || ||| |        || ||||| || |||||||| ||||| ||
Db        361 GTGCGTAAAGATACTCGAGGCTTAATAGCGCGTTGGAAATATTTCTGGCAATCGGTATTA 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  | |  |  ||  |   |  ||| || ||  ||  |   |  ||   ||
Db        421 GCTTTAGCGGTTGCATTTAGTATGTACGCGATTGGTAAAGATACCCCAGCAACGCAACTT 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              |||||||  |  | ||   |||  |   ||  | ||||| || |   |  | ||| |
Db        481 GTCGTGCCTTTCTTTAAAGATGTCATGCCACAACTTGGCATGTTGTATATCTTACTGGCT 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              || || |||||  || ||   | || || |||||  ||  | |||||  ||| | || ||| |
Db        541 TATTTTGTCATTGTTGGTACCAGTAACGCGGTAAATTTAACCGATGGTTTAGATGGCTTA 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              || || |||||  || ||  | || || |  |  | |  |  |  |  |  | ||
Db        601 GCCATTATGCCTACCGTGTTTGTAGCGGCTGGATTTGCATTAGTTGCATGGGCTACAGGT 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              || ||  || ||  |||   |||||    |||||   |  ||  |   || |||||  || |
Db        661 AATGTCAATTTTGCCAGCTACCTAAAAATTCCTTATTTAATGCATGCAGGCGAATTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  | ||  ||||  | ||||| || ||  | |||||||| ||||| || || |||||
Db        721 ATTGTTTGTACCGCCATTGTCGGTGCGGGATTAGGCTTCCTTTGGTTTAATACTTATCCT 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCCGCGCTGGGCACCATC 840
              || || ||||| ||||| || || ||| |  | ||  | || | |||  | || || ||| |
Db        781 GCCCAAGTCTTTATGGGTGATGTTGGCTCTTTAGCATTAGGTGGCGCATTAGGTACTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  | ||||| ||  |  |  |  | || ||||| || ||||||||  | |||
Db        841 GCCGTGTTATTACGCCAAGAATTTTTATTAGTGATTATGGGGGGCGTGTTCGTTGTTGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || ||||||| ||  | || || | ||| || || |||  ||| |  ||  | ||||||
Db        901 ACGCTCTCGGTTATTTTACAAGTAGGTTCATTTAAACTGCGCGGTCAACGTATTTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||| |||||||| || |  ||| | ||||| ||||| || || || ||||| ||
Db        961 ATGGCGCCAATCCATCATCACTATGAATTAAAAGGTTGGCCAGAGCCTCGTGTGATTGTC 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||| |||||||| |  | ||  | || || || ||  | || || || ||| | |||
Db       1021 CGCTTTTGGATCATCTCTCTTATGTTAGTCCTTATTGGTTTAGCGACATTAAAGGTACGT 1080

Qy       1081 TGA 1083
              | |
Db       1081 TAA 1083
```

```
RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
;   GENERAL INFORMATION:
;        APPLICANT: Fleischmann et al.
;        TITLE OF INVENTION: The Nucleotide sequence of
;                            the Haemophilus influenzae Rd Genome, Fragments
;                            Thereof, and Uses Thereof
;        NUMBER OF SEQUENCES: 1
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: MD
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 3 1/2 inch diskette
;             COMPUTER: Dell Pentium
;             OPERATING SYSTEM: MS DOS v6.22
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/557,884
;             FILING DATE: 25-Apr-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/476,102
;             FILING DATE: JUN-5-1995
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Michelle S. Marks
;             REGISTRATION NUMBER: 41,971
;             REFERENCE/DOCKET NUMBER: PB186P3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 301-309-8504
;             TELEFAX: 301-309-8439
;    INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1830121 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

  Query Match             33.4%;  Score 361.4;  DB 4;  Length 1830121;
  Best Local Similarity   58.4%;  Pred. No. 4.6e-72;
  Matches  632;  Conservative    0;  Mismatches  451;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              |||  |   |||  || |||||  ||      | | | | |  ||    |  |
Db    1201927 ATGTTAGTCTGGCTTGCTGAATATCTTGTTCGTTACGAAACCGCGTTTAATGCTATTTCT 1201986

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||  | ||  | ||||  | |||      | |||||  ||  |   | || || ||| |
Db    1201987 TATATTACAGTCCGCGCAAATCTTGCATTATTAACCGCACTTTTTATCTCACTTTGGATT 1202046

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||   |||||    |||||||||      | ||||| |  |||| || ||
Db    1202047 GGCCCTAAAGTGATCAAACGCTTGCAGATCTTAAAATTTGGCCAAGAAGTGCGAAATGAT 1202106

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||  |   ||| |  | || ||  ||||| || || ||||| || |   ||||  |
Db    1202107 GGCCCTGAAAGTCACTTTGCAAAAAAAGGCACACCCACTATGGGTGGTGTGATGATTTTA 1202166

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              | | || | | | || ||| |  | |||||  |||||  | || |  ||  | |||||
Db    1202167 TTCTCTATTGGCGTAAGTACGTTATTATGGGCAAATCTTGCTAATCCGTATATTTGGGTT 1202226

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |  | |||      |    | ||| || || ||  || || || | ||| ||  |
Db    1202227 TGTTTATTTGTTTTATTTGGATACGGCGCAATTGGTTTTGTGGATGATTTCCGTAAAATT 1202286

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              | |   || ||  |  |||  ||       || ||||| || ||||||  ||| ||| |
Db    1202287 ACCCGTAAAAATACTGATGGATTGATTGCTCGTTGGAAATATTTCTGGATGTCTGTGGTG 1202346

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  | |  ||   | | |  || ||   ||   | | |   |  |||   |
Db    1202347 GCATTAGTGGCAATCCTTTGGCTTTATTGGCTTGGTCACGACACTGATGCCACCCGTTTA 1202406

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              |  | ||  |  | ||   | |   |  || || ||  | ||||   | ||  |  |
Db    1202407 GTGATTCCATTCTTTAAAGACATTATGCCTCAATTAGGTTTGTTCTATATTGTGTTATCT 1202466

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||| || || ||| ||  | |||||  |||||   | |||||||| | || ||  |
Db    1202467 TACTTTGTGATTGTTGGCACGGGTAATGCAGTGAATTTAACCGACGGTTTAGATGGATTA 1202526
```

```
Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||| ||||| || |    | ||||| || ||| | |    |  |      |      | ||
Db    1202527 GCGATTATGCCTACTGCGCTTGTTGCAGGTGCGTTTGCTTTAATTGCTTGGGCTACAGGT 1202586

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              |||||||| ||||| || ||  ||| |||||  | |       |       ||  | |
Db    1202587 AACGTGAATTTCGCAGAATATTTGCACATTCCGTATATTAAATACAGTTCTGAAGTAGTG 1202646

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |||||||  | ||  | || |||||  |  | || ||| | ||||| || || |||||
Db    1202647 GTGTTCTGTACAGCTATTGTTGGCGCGAGTTTGGGATTCTTATGGTTTAATACTTATCCA 1202706

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              || || || || ||||| ||||| ||  | || ||  | || |  ||||||||     |
Db    1202707 GCTCAAGTATTTATGGGGGACGTGGGTTCTCTAGCATTAGGTGGTGCGCTGGGTGTTGTA 1202766

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||  |  | || || |||||  |  ||||  | || |||||||| || || ||  | |||
Db    1202767 GCAATTCTTGTTCGTCAGGAATTTTTGCTTGTGATAATGGGTGGTGTATTTGTTGTTGAA 1202826

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              | ||||| || ||  | || ||  | ||  |  ||  |  ||  ||| |  ||  |||  |
Db    1202827 GCACTCTCTGTTATTTTGCAAGTAGGGTCTTATAAGTTACGCAAACAACGCATTTTTAGA 1202886

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              ||||| ||||| || |||||||| ||| |||| || ||||| || ||  | |||||  |
Db    1202887 ATGGCACCGATTCACCACCATTTTGAATTGAAGGGATGGCCTGAGCCAAGAGTGATTATT 1202946

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              || || ||||| ||  || | || |||||| |||| ||  | | |||||||||| |||||
Db    1202947 CGGTTTTGGATTATTTCCTTAATGCTGGTGTTGATGGGATTGGTCACCTTGAAGTTGCGT 1203006

Qy       1081 TGA 1083
              | |
Db    1203007 TAA 1203009
```

RESULT 6
US-09-643-990A-1
; Sequence 1, Application US/09643990A
; Patent No. 6528289
; GENERAL INFORMATION:
; APPLICANT: Robert D. Fleischmann
; Mark D. Adams
; Owen White
; Hamilton O. Smith
; J. Craig Venter
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville,
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/643,990A
; FILING DATE: 23-Aug-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: 1995-06-07
; APPLICATION NUMBER: 08/426,787
; FILING DATE: 1995-04-21
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB186P1C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-610-5790
; TELEFAX: 310-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 33.4%; Score 361.4; DB 4; Length 1830121;
Best Local Similarity 58.4%; Pred. No. 4.6e-72;
Matches 632; Conservative 0; Mismatches 451; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |   ||| || |||||  ||    | | | |  |  ||    |  |
Db    1201927 ATGTTAGTCTGGCTTGCTGAATATCTTGTTCGTTACGAAACCGCGTTTAATGCTATTTCT 1201986

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||  | ||  | |||| | |||       | | |||||  || |   | || ||| | |
Db    1201987 TATATTACAGTCCGCGCAAATCTTGCATTATTAACCGCACTTTTTATCTCACTTTGGATT 1202046

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||     |||||     |||||||||     |  | ||||| |  ||||| || ||
Db    1202047 GGCCCTAAAGTGATCAAACGCTTGCAGATCTTAAAATTTGGCCAAGAAGTGCGAAATGAT 1202106

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||  |    ||| |  || || ||||| || || ||||| || |   |||| |
Db    1202107 GGCCCTGAAAGTCACTTTGCAAAAAAAGGCACACCCACTATGGGTGGTGTGATGATTTTA 1202166

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              | | || | | | || ||| |  | ||||  ||||| | ||  ||  | |||||
Db    1202167 TTCTCTATTGGCGTAAGTACGTTATTATGGGCAAATCTTGCTAATCCGTATATTTGGGTT 1202226

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |  | |||    |  | |||  || || ||  ||  ||  | ||| ||  |
Db    1202227 TGTTTATTTGTTTTATTTGGATACGGCGCAATTGGTTTTGTGGATGATTTCCGTAAAATT 1202286

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              | |   || ||  |  ||| ||        || ||||| || ||||| ||| ||| |
Db    1202287 ACCCGTAAAAATACTGATGGATTGATTGCTCGTTGGAAATATTTCTGGATGTCTGTGGTG 1202346

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  | |  ||  | | |  || ||  |  ||  |  | |  |  |||  |
Db    1202347 GCATTAGTGGCAATCCTTTGGCTTTATTGGCTTGGTCACGACACTGATGCCACCCGTTTA 1202406

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              |  | ||  |  | ||  | |  |  || || ||  | |||  | ||  | |
Db    1202407 GTGATTCCATTCTTTAAAGACATTATGCCTCAATTAGGTTTGTTCTATATTGTGTTATCT 1202466

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||| || || || ||| || | ||||| |||||  | |||||||||| | || ||  |
Db    1202467 TACTTTGTGATTGTTGGCACGGGTAATGCAGTGAATTTAACCGACGGTTTAGATGGATTA 1202526

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||| ||||| || |    | ||||| || ||| | |    |  |      |      | ||
Db    1202527 GCGATTATGCCTACTGCGCTTGTTGCAGGTGCGTTTGCTTTAATTGCTTGGGCTACAGGT 1202586

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              |||||||| ||||| || ||  ||| |||||  | |       |       ||  | |
Db    1202587 AACGTGAATTTCGCAGAATATTTGCACATTCCGTATATTAAATACAGTTCTGAAGTAGTG 1202646

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |||||||  | ||  | || |||||  |  | || ||| | ||||| || || |||||
Db    1202647 GTGTTCTGTACAGCTATTGTTGGCGCGAGTTTGGGATTCTTATGGTTTAATACTTATCCA 1202706

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
```

```
Db    1202707 GCTCAAGTATTTATGGGGGACGTGGGTTCTCTAGCATTAGGTGGTGCGCTGGGTGTTGTA 1202766

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900

Db    1202767 GCAATTCTTGTTCGTCAGGAATTTTTGCTTGTGATAATGGGTGGTGTATTTGTTGTTGAA 1202826

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960

Db    1202827 GCACTCTCTGTTATTTTGCAAGTAGGGTCTTATAAGTTACGCAAACAACGCATTTTTAGA 1202886

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020

Db    1202887 ATGGCACCGATTCACCACCATTTTGAATTGAAGGGATGGCCTGAGCCAAGAGTGATTATT 1202946

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080

Db    1202947 CGGTTTTGGATTATTTCCTTAATGCTGGTGTTGATGGGATTGGTCACCTTGAAGTTGCGT 1203006

Qy       1081 TGA 1083

Db    1203007 TAA 1203009
```

RESULT 7
US-09-328-352-2043
; Sequence 2043, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 2043
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-2043

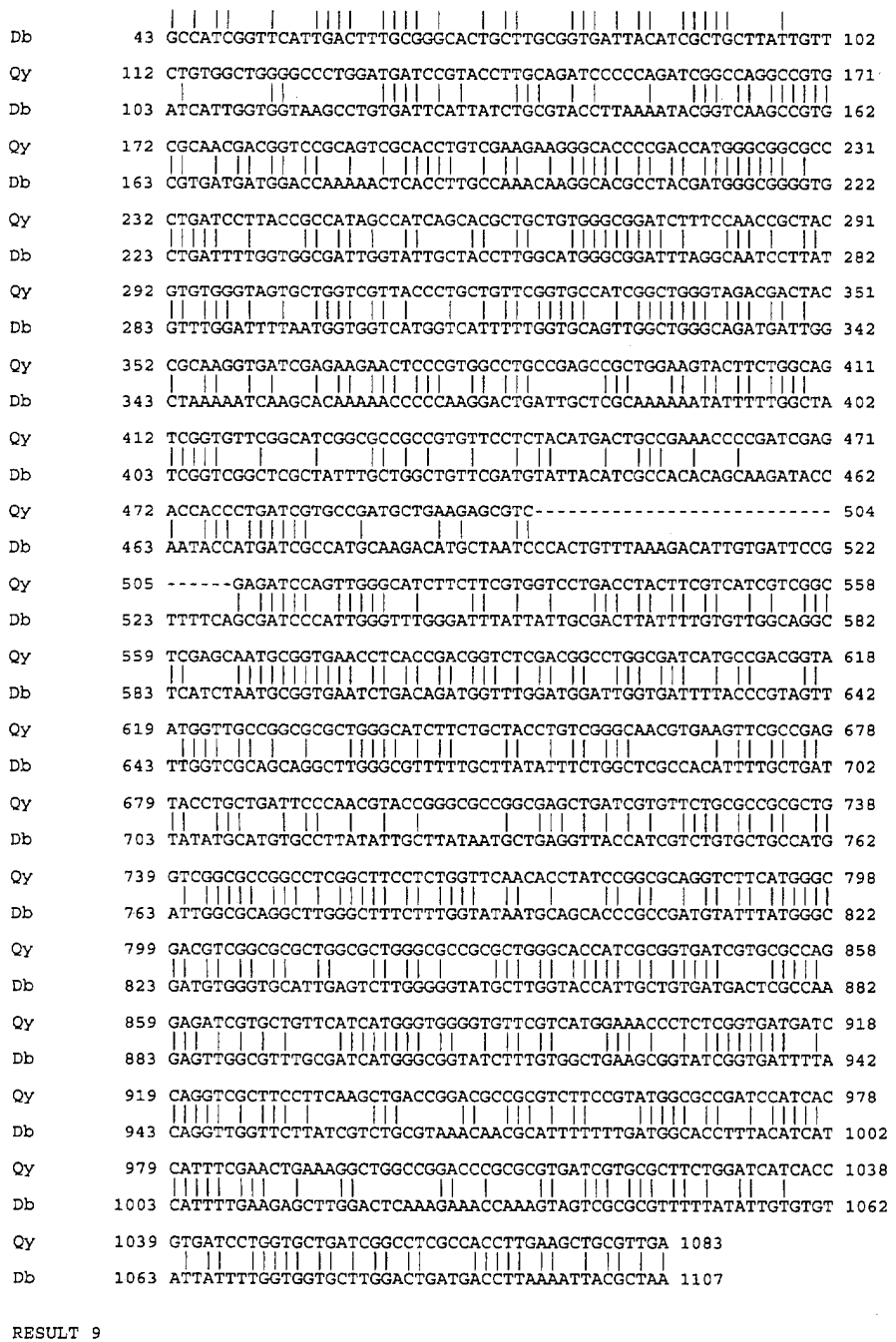
Query Match 23.1%; Score 250.6; DB 4; Length 1119;
Best Local Similarity 53.4%; Pred. No. 8.7e-48;
Matches 597; Conservative 0; Mismatches 484; Indels 36; Gaps 2;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60

Db          1 ATGCTGTTATGGTTGTTTGAACAACTTGCGGGCTATCACAGTTCGTTTCAGGTTGTTCGT 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120

Db         61 TATTTAACATTACGTTCTTTACTCAGTGTATTAACGTCACTGACCATTGGTTTGGTTCTC 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180

Db        121 GGACCAATCATGATTCGTAAATTACAAGCGTTAAAATACGGTCAGGCAGTAAGTTCGTTT 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240

Db        181 GCTCCTGAAAATCATGCTAAGAAGATGGGTACACCGACCATGGGTGGGATTTTAATTCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300

Db        241 CTCTCAATTGGTATTAGTACTTTATTGTGGGCTGATTTATCTAACCCTTATGTTTGGATT 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360

Db        301 GTACTTGGTGTAATGGTTGTGTTCGGTGCGGTTGGCTGGGCAGATGACTGGATTAAAATT 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420

Db        361 CGCTATAAAGATAATGCAGGTCTACCTGCCCGTAAGAAGTTTTTCTGGACTTCTGTCGCA 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTG----CCGAAACCCCGATCGAGACCAC 476

Db        421 TCGCTGGGTGCTGGTATTGCGTTATATTTAATTGCTACCCAACAATCTAATGCAGAATAC 480

Qy        477 CCTGATCGTGCCGATGCTGAAGAGCGTCGAG---------------------------- 507

Db        481 ACAGCAAATATGCTGGATTTATTGATTCCTTTCTTTAAGAATCTTTCAATTCCGCTTTCT 540

Qy        508 ---ATCCAGTTGGGCATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGC 564

Db        541 ATTGTTCCTTTAGGTTTAGCATTTATTGTATTTACCTATTTAGTAATTAATGGTGCCTCT 600

Qy        565 AATGCGGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTT 624

Db        601 AATGCGGTTAACCTAACTGATGGTCTAGATGGTTTAGCCATTATGCCAGTAGTTATGGTA 660

Qy        625 GCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTG 684

Db        661 GCGACGGGCTTAGGCGTATTTGCGTATTTATCTGGTGATATTCGTTTTGCCAACTATTTG 720

Qy        685 CTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGC 744

Db        721 CATATTCCTTATGTGAAATACACTTCAGAACTTGTGGTGATCTGTTCTGCCATGATTGGT 780

Qy        745 GCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTC 804

Db        781 GCGGGTTTGGCATTTCTTTGGTATAACGCACATCCGGCGCAAGTCTTTATGGGTGATGTG 840

Qy        805 GGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATC 864

Db        841 GGCGCTTTAGCTCTAGGTGCAATGCTTGGAACCATTGCCGTTATGGTTCGCCAAGAAATC 900

Qy        865 GTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTC 924

Db        901 GTATTCGCAATTATGGGCGGTGTTTTCGTTATGGAAGCGGTATCGGTATTCCTGCAAATC 960

Qy        925 GCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTC 984

Db        961 GGCTCATTACGAATGCGGAATAAACGCGTATTTTTGATGGCACCTTTGCACCATCACTAT 1020

Qy        985 GAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATC 1044

Db       1021 GAAAAACAAGGCTGGAAAGAAACCCAAGTAGTTATCCGTTTCTGGATTATTACGATTATG 1080

Qy       1045 CTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGTT 1081

Db       1081 CTGGTTGTTTTAGGCCTAATGACCTTAAAATTGCGCT 1117
```

RESULT 8
US-09-540-236-1218
; Sequence 1218, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1218
; LENGTH: 1107
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-1218

Query Match 21.6%; Score 233.8; DB 4; Length 1107;
Best Local Similarity 52.6%; Pred. No. 5.4e-44;
Matches 560; Conservative 0; Mismatches 472; Indels 33; Gaps 1;

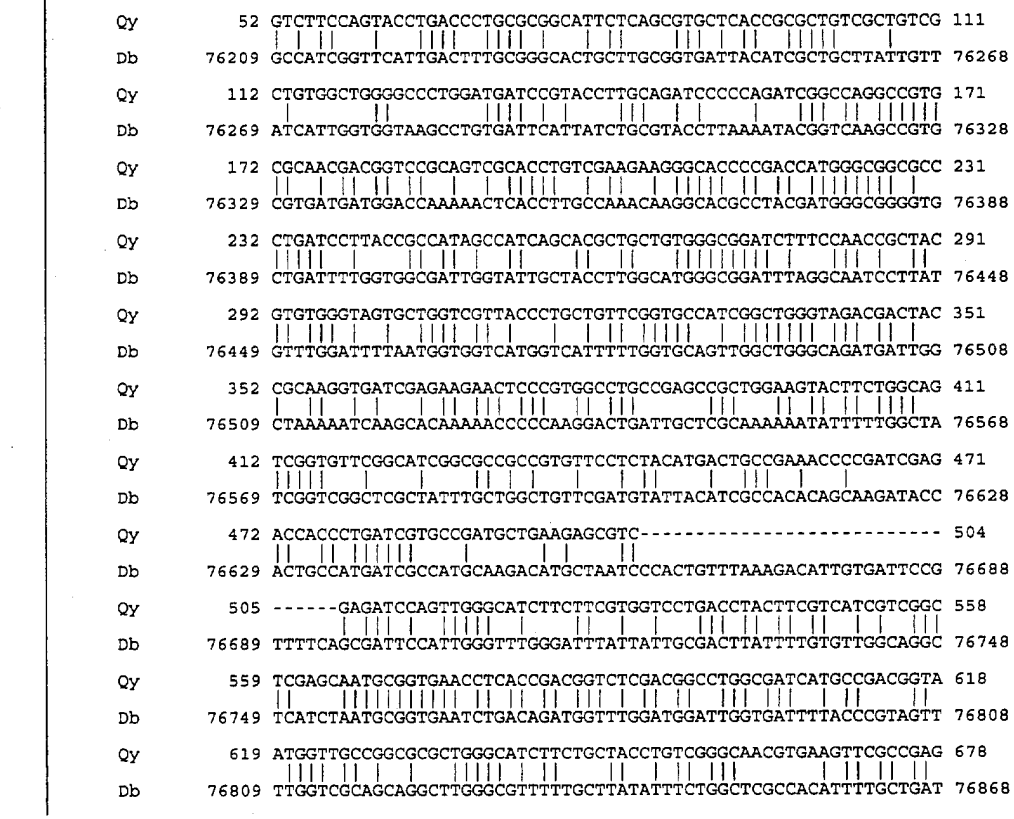
```
Qy         52 GTCTTCCAGTACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCG 111
```



RESULT 9

US-09-596-002-39
; Sequence 39, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 100848
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 39
; PUBLICATION INFORMATION:
US-09-596-002-39

Query Match 21.4%; Score 232.2; DB 4; Length 100848;
Best Local Similarity 52.5%; Pred. No. 3.4e-43;
Matches 559; Conservative 0; Mismatches 473; Indels 33; Gaps 1;



```
Qy        679 TACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTG 738
Db      76869 TATATGCATGTGCCTTATATTGCTTATAATGCTGAGGTTACCATCGTCTGTGCTGCCATG 76928

Qy        739 GTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGC 798
Db      76929 ATTGGCGCAGGCTTGGGCTTTCTTTGGTATAATGCAGCACCCGCCGATGTATTTATGGGC 76988

Qy        799 GACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAG 858
Db      76989 GATGTGGGTGCATTGAGTCTTGGGGGTATGCTTGGTACCATTGCTGTGATGACTCGCCAA 77048

Qy        859 GAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATC 918
Db      77049 GAGTTGGCGTTTGCGATCATGGGCGGTATCTTTGTGGCTGAAGCGGTATCGGTGATTTTA 77108

Qy        919 CAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGTATGGCGCCGATCCATCAC 978
Db      77109 CAGGTTGGTTCTTATCGTCTGCGTAAACAACGCATTTTTTTGATGGCACCTTTACATCAT 77168

Qy        979 CATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATCATCACC 1038
Db      77169 CATTTTGAAGAGCTTGGACTCAAAGAAACCAAAGTAGTCGCGCGTTTTTATATTGTGTGT 77228

Qy       1039 GTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGTTGA 1083
Db      77229 ATTATTTTGGTGGTGCTTGGACTGATGACCTTAAAATTACGCTAA 77273
```

RESULT 10
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 16.2%; Score 175.4; DB 3; Length 4403765;
Best Local Similarity 50.9%; Pred. No. 5.3e-30;
Matches 527; Conservative 0; Mismatches 466; Indels 42; Gaps 3;

```
Qy          78 CATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTGGGGCCCTGGATGATCCG 137
Db     2416121 CCTTATCGCCGTTGCCGTAGCGGTGACGGTGTCCATCTTGCTGACCCCGGTGCTGATCCG 2416062

Qy         138 TACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGACGGTCCGCAGTCGCACCT 197
Db     2416061 GTTGTTCACTAAGCAGGGCTTCGGCCACCAGATCCGTGAGGATGGCCCGCCCAGCCACCA 2416002

Qy         198 GTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTTACCGCCATAGCCATCAG 257
Db     2416001 CACCAAGCGCGGTACGCCGTCGATGGGCGGGGTGGCGATTCTGGCCGGCATCTGGGCGGG 2415942

Qy         258 CACGCTGC------------------------TGTGGGCGGATCTTTCCAACCGCTACGT 293
```

```
Db     2415941 CTACCTGGGCGCCCACCTAGCGGGCCTGGCGTTTGACGGTGAAGGCATCGGCGCATCGGG 2415882

Qy         294 GTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCG 353
Db     2415881 TCTGTTGGTGCTGGGCCTAGCCACCGCTTTGGGCGGCGTCGGGTTCATCGACGATCTGAT 2415822

Qy         354 CAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAG-- 411
Db     2415821 CAAGATCCGCAGGTCGCGCAATCTCGGGTTGAACAAGACGGCCAAGACCGTCGGGCAGAT 2415762

Qy         412 ----------TCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAAC 461
Db     2415761 CACCTCCGCCGTGCTGTTTGGCGTGCTGGTGCTGCAGTTCCGGAATGCTGCCGGCCTGAC 2415702

Qy         462 CCCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCAT 521
Db     2415701 ACCGGGCAGCGCGGATCTGTCCTACGTGCGTGAGATCGCCACCGTCACATTGGCGCCGGT 2415642

Qy         522 CTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCAC 581
Db     2415641 GCTGTTCGTGTTGTTCTGCGTGGTCATCGTCAGCGCCTGGTCGAACGCGGTCAACTTCAC 2415582

Qy         582 CGACGGTCTCGACGGCCTGGC------GATCATGCCGACGGTAATGGTTGCCGGCGCGCT 635
Db     2415581 CGATGGCCTGGACGGGCTGGCCGCCGGCACCATGGCGATGGTCACCGCCGCCTACGTGCT 2415522

Qy         636 GGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAA 695
Db     2415521 GATCACCTTCTGGCAGTACCGCAACGCGTGCGTGACGGCGCCGGGCCTGGGCTGCTACAA 2415462

Qy         696 CGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGG 755
Db     2415461 CGTGCGCGACCCGCTGGACCTGGCGCTCATCGCGGCCGCAACCGCTGGCGCCTGCATCGG 2415402

Qy         756 CTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGC 815
Db     2415401 TTTTTTGTGGTGGAACGCCGCGCCCGCCAAGATCTTCATGGGTGACACTGGGTCGCTGGC 2415342

Qy         816 GCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCAT 875
Db     2415341 GTTGGGCGGCGTCATCGCGGGGTTGTCGGTGACCAGCCGCACCGAGATCCTTGCGGTGGT 2415282

Qy         876 CATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAA 935
Db     2415281 GCTGGGTGCGCTGTTCGTCGCCGAGATCACCTCGGTGGTGTTGCAAATCCTGACCTTCCG 2415222

Qy         936 GCTGACCGGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGG 995
Db     2415221 GACCACCGGGCGCCGGATGTTTCGGATGGCGCCCTTCCACCACCATTTCGAGTTGGTCGG 2415162

Qy         996 CTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGAT 1055
Db     2415161 TTGGGCTGAAACCACGGTCATCATCCGGTTCTGGCTGCTCACCGCGATCACCTGCGGTCT 2415102

Qy        1056 CGGCCTCGCCACCTT 1070
Db     2415101 GGGCGTGGCCTTGTT 2415087
```

RESULT 11
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

```
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4411529
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   OTHER INFORMATION: H37Rv
US-09-103-840A-1

  Query Match            16.2%;  Score 175.4;  DB 3;  Length 4411529;
  Best Local Similarity   50.9%;  Pred. No. 5.3e-30;
  Matches  527;  Conservative    0;  Mismatches  466;  Indels   42;  Gaps    3;

Qy         78 CATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTGGGGCCCTGGATGATCCG 137
Db    2417462 CCTTATCGCCGTTGCCGTAGCGGTGACGGTGTCCATCTTGCTGACCCCGGTGCTGATCCG 2417403

Qy        138 TACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGACGGTCCGCAGTCGCACCT 197
Db    2417402 GTTGTTCACTAAGCAGGGCTTCGGCCACCAGATCCGTGAGGATGGCCCGCCCAGCCACCA 2417343

Qy        198 GTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTTACCGCCATAGCCATCAG 257
Db    2417342 CACCAAGCGCGGTACGCCGTCGATGGGCGGGGTGGCGATTCTGGCCGGCATCTGGGCGGG 2417283

Qy        258 CACGCTGC----------------------------TGTGGGCGGATCTTTCCAACCGCTACGT 293
Db    2417282 CTACCTGGGCGCCCACCTAGCGGGCCTGGCGTTTGACGGTGAAGGCATCGGCGCATCGGG 2417223

Qy        294 GTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCG 353
Db    2417222 TCTGTTGGTGCTGGGCCTAGCCACCGCTTTGGGCGGCGTCGGGTTCATCGACGATCTGAT 2417163

Qy        354 CAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAG-- 411
Db    2417162 CAAGATCCGCAGGTCGCGCAATCTCGGGTTGAACAAGACGGCCAAGACCGTCGGGCAGAT 2417103

Qy        412 ----------TCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAAC 461
Db    2417102 CACCTCCGCCGTGCTGTTTGGCGTGCTGGTGCTGCAGTTCCGGAATGCTGCCGGCCTGAC 2417043

Qy        462 CCCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCAT 521
Db    2417042 ACCGGGCAGCGCGGATCTGTCCTACGTGCGTGAGATCGCCACCGTCACATTGGCGCCGGT 2416983

Qy        522 CTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCAC 581
Db    2416982 GCTGTTCGTGTTGTTCTGCGTGGTCATCGTCAGCGCCTGGTCGAACGCGGTCAACTTCAC 2416923

Qy        582 CGACGGTCTCGACGGCCTGGC------GATCATGCCGACGGTAATGGTTGCCGGCGCGCT 635
Db    2416922 CGATGGCCTGGACGGGCTGGCCGCCGGCACCATGGCGATGGTCACCGCCGCCTACGTGCT 2416863

Qy        636 GGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAA 695
Db    2416862 GATCACCTTCTGGCAGTACCGCAACGCGTGCGTGACGGCGCCGGGCCTGGGCTGCTACAA 2416803

Qy        696 CGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGG 755
Db    2416802 CGTGCGCGACCCGCTGGACCTGGCGCTCATCGCGGCCGCAACCGCTGGCGCCTGCATCGG 2416743

Qy        756 CTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGC 815
Db    2416742 TTTTTTGTGGTGGAACGCCGCGCCCGCCAAGATCTTCATGGGTGACACTGGGTCGCTGGC 2416683

Qy        816 GCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCAT 875
Db    2416682 GTTGGGCGGCGTCATCGCGGGGTTGTCGGTGACCAGCCGCACCGAGATCCTTGCGGTGGT 2416623

Qy        876 CATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAA 935
Db    2416622 GCTGGGTGCGCTGTTCGTCGCCGAGATCACCTCGGTGGTGTTGCAAATCCTGACCTTCCG 2416563

Qy        936 GCTGACCGGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGG 995
Db    2416562 GACCACCGGGCGCCGGATGTTTCGGATGGCGCCCTTCCACCACCATTTCGAGTTGGTCGG 2416503

Qy        996 CTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGAT 1055
Db    2416502 TTGGGCTGAAACCACGGTCATCATCCGGTTCTGGCTGCTCACCGCGATCACCTGCGGTCT 2416443

Qy       1056 CGGCCTCGCCACCTT 1070
Db    2416442 GGGCGTGGCCTTGTT 2416428


RESULT 12
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
;  APPLICANT: SHIGENOBU, SHUJI
;  APPLICANT:  WATANABE, HIDEMI
;  APPLICANT:  HATTORI, MASAHIRA
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
;  FILE REFERENCE: 081356/0159
;  CURRENT APPLICATION NUMBER: US/09/790,988
;  CURRENT FILING DATE:  2001-02-23
;  PRIOR APPLICATION NUMBER: JP2000-107160
;  PRIOR FILING DATE: 2000-04-07
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 640681
;   TYPE: DNA
;   ORGANISM: Buchnera sp.
US-09-790-988-1

  Query Match            16.1%;  Score 174.2;  DB 4;  Length 640681;
  Best Local Similarity   48.3%;  Pred. No. 6.4e-30;
  Matches  488;  Conservative    0;  Mismatches  523;  Indels    0;  Gaps    0;

Qy         73 CGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTGGGGCCCTGGATG 132
Db    242303 CGTGCAATATTTAGCTTATTAACTTCTTTTTTATCAATCTATATATTGGGCCATATTTT 242244

Qy        133 ATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGACGGTCCGCAGTCG 192
Db    242243 ATATATTACTTTAAAAAATTACAAACATATCAAATTATACGAAATAATGGGCCAAAAACA 242184

Qy        193 CACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTTACCGCCATAGCC 252
Db    242183 CATTATTCAAAAAAAAACACACCTACCATGGGTGGAATTTTTATAATTTTTTCTATATTG 242124

Qy        253 ATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTAGTGCTGGTCGTT 312
Db    242123 TTTTCGACAATACTATACTGTAATTTATCCAATATTTATATTTGGTACGTCATAAGTATT 242064

Qy        313 ACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTGATCGAGAAGAAC 372
Db    242063 TTAATAGGATATGGCTTAATAGGTTTTATAGATGATTATAAAAAAATAAAGTATAAAAAT 242004

Qy        373 TCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTCGGCATCGGCGCC 432
Db    242003 AGTCAAGGATTAAAACTAAAATGGAAATATTTTTTTTTATCTATTATTGCTTTCATATTC 241944

Qy        433 GCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTGATCGTGCCGATG 492
Db    241943 ATCTGTATGATTAAGATAAACAATAAAGATATTATTTCTACTGAATTAATAATTCCCTTT 241884

Qy        493 CTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACCTACTTCGTCATC 552
Db    241883 TGTATAAAGAATGATTTTGAAATAAACTATTTATATATATTTCTTTCTTATTTTGTACTA 241824

Qy        553 GTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTGGCGATCATGCCG 612
```

```
Db     241823 GTTGGCACTAGTAATGCAGTTAATCTAACAGATGGATTAGATGGTTTAGCTATTATGCCA 241764

Qy        613 ACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTC 672

Db     241763 GTTATTTTTTTAACTTGTGGTTTGACATTAATATCTTTATTTAGCGACAATATAAATATC 241704

Qy        673 GCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCC 732

Db     241703 TCACATTATCTACATGTTCAATATGTAAAAAATTCAACAGAACTAGCTATATTATGTATG 241644

Qy        733 GCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTC 792

Db     241643 GCAATTGTTGGTTCAGGATTAGGTTTTTTATGGTTTAATAGCTATCCAGCAAAAGTGTTC 241584

Qy        793 ATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTG 852

Db     241583 ATGGGAGATGTTGGATCACTAGCATTAGGTGGTTCTTTAGGAGCTATAGCAATATTATTA 241524

Qy        853 CGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTG 912

Db     241523 CATCAAGAATTACTACTAATTATCATGGGTGGTATTTTTGTTTTTGAAACAATATCTGTG 241464

Qy        913 ATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGTATGGCGCCGATC 972

Db     241463 ATATTACAAATTATATCGTTTAAAATAAGAAAAAAAAGAATTTTTCAAATGGCACCAGTT 241404

Qy        973 CATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATC 1032

Db     241403 CACCACCATTATGAAGTCAAAGGAATATTAGAACCTTTGATTATCGTAAGATTTTGGATT 241344

Qy       1033 ATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGTTGA 1083

Db     241343 GTGTCACTAATATTGTTACTCATAAGTCTCATTTCTTTAAAGGTATGTTGA 241293


RESULT 13
US-09-489-039A-5281
; Sequence 5281, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 5281
;   LENGTH: 210
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5281

  Query Match           12.4%;  Score 134;  DB 4;  Length 210;
  Best Local Similarity   80.5%;  Pred. No. 1.2e-21;
  Matches  169;  Conservative    0;  Mismatches   40;  Indels    1;  Gaps    1;

Qy        875 TCATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCA 934

Db          1 TCATGGGCGGGGTTTTCGTGGTGGAAACGCTGTCGGTTATTCTGCAGGTCGGCTCCTTTA 60

Qy        935 AGCTG-ACCGGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAA 993

Db         61 AGCTGCGCCGGTCAGCGCATCTTCCGCATGGCGCCGATTCACCACCACTATGAACTGAAA 120

Qy        994 GGCTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTG 1053

Db        121 GGCTGGCCGGAGCCGCGCGTTATCGTGCGCTTCTGGATTATTTCGCTGATGCTGGTGCTG 180

Qy       1054 ATCGGCCTCGCCACCTTGAAGCTGCGTTGA 1083

Db        181 ATTGGCCTGGCGACGCTGAAGGTACGTTAA 210


RESULT 14
US-08-311-174-4
; Sequence 4, Application US/08311174
; Patent No. 5556776
; GENERAL INFORMATION:
;    APPLICANT:  TSUCHIYA, MAKOTO
;    APPLICANT:  MIWA, KIYOSHI
;    TITLE OF INVENTION:  SUCRASE GENE DERIVED FROM CORYNEFORM
;    TITLE OF INVENTION:  BACTERIA
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/311,174
;      FILING DATE:  23-SEP-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 046836/1992
;      FILING DATE:  04-MAR-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Oblon, No. 5556776man F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  10-699-0
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6911 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
US-08-311-174-4

  Query Match           12.3%;  Score 133.4;  DB 1;  Length 6911;
  Best Local Similarity   54.9%;  Pred. No. 3.7e-21;
  Matches  308;  Conservative    0;  Mismatches  246;  Indels    7;  Gaps    2;

Qy        519 CATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCT 578

Db       6079 CATCGTGTTCCTCATTTTTATCTACGTTGTGGTCAGCGCGTGGTCGAATGCCGTGAACAT 6138

Qy        579 CACCGACGGTCTCGACGGCCTGGCG------ATCATGCCGACGGTAATGGTTGCCGGCGC 632

Db       6139 CACTGATGGTTTGGATGGTTTGGCTGCAGGTACCACAGCATTTGTCATGGGTGCTTACAC 6198

Qy        633 GCTGGGCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCC 692

Db       6199 CTTGATCACGTTCTGGCAGTTCCGAAACTCCTGCGATACTGCAGTGGAAGCGGGTTGCAA 6258

Qy        693 CAACGTACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCT 752

Db       6259 TACGGTGCGTGATCCACTGGATTTGTCTGTGTTGTGCGCTGC-TGGTCTGGCGCCACCTT 6317

Qy        753 CGGCTTCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCT 812
```

```
Db       6318 GGGCTTTCTGTGGTGGAATGCGGCACCGACAAAGATCTTCATGGGCGATACTGGTTCTTT 6377

Qy        813 GGCGCTGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTT 872

Db       6378 GGCACTGGGCGGTTTGGTTGCAGGTATTTCTGTGGTTAGCCGCACCGAGCTGCTCATGGT 6437

Qy        873 CATCATGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTT 932

Db       6438 TATCATCGGCGCGCTGTTTGTCATTGAGGTCGCTTCTGTTGCGATCCAGATCGGCGTGTT 6497

Qy        933 CAAGCTGACCGGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAA 992

Db       6498 TAAGACCCGCGGTAAGCGTGTGTTCAAAATGGCTCCGATCCACCACCACTTCGAGGCCCT 6557

Qy        993 AGGCTGGCCGGACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCT 1052

Db       6558 TGGGTGGACTGAAACTACCGTGACCATCCGTTTCTGGCTGATCACGATCATGACTGTGTT 6617

Qy       1053 GATCGGCCTCGCCACCTTGAA 1073

Db       6618 GGCGGGTGTCGGTGTGTTTAA 6638


RESULT 15
US-09-221-017B-881
; Sequence 881, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  881:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2402 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...2402
US-09-221-017B-881

  Query Match             10.6%;  Score 115.2;  DB 4;  Length 2402;
  Best Local Similarity   54.5%;  Pred. No. 3.8e-17;
  Matches  231;  Conservative    0;  Mismatches  193;  Indels    0;  Gaps    0;

Qy        518 GCATCTTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACC 577

Db        500 GGATTCTCTTCGTCTGCGTAGCTGTAGTGGTAGTCACCTTTATATCCAACTGTGCCAATC 559

Qy        578 TCACCGACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGG 637

Db        560 TGACTGATGGGTTGGATGGTTTAGCGACGGGGTCTTCTGCTATTATCGGAGTTGTACTGG 619

Qy        638 GCATCTTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACG 697

Db        620 CTATTTTTGCCTATGTCTCCTCCCATATCGAGATGGCCTCATATCTGAATATCATGTTTA 679

Qy        698 TACCGGGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCT 757

Db        680 TTCCCGGAGCAGAGGAGCTGACCATTTTTGCTTTTGCTTTCGTTGGTGCCACGATCGGTT 739

Qy        758 TCCTCTGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGC 817

Db        740 TTCTGTGGTACAATGCTTATCCGGCACAGGTCTTTATGGGAGATACAGGAAGTCTCACTT 799

Qy        818 TGGGCGCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCA 877

Db        800 TGGGAGGAATCATCGCTGTTTTCGCTCTGATCATTCGCAAAGAAATGCTACTGCCCATCC 859

Qy        878 TGGGTGGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGC 937

Db        860 TCTGCTTCGTATTTATCATAGAGGGTCTTTCCGTGATGATTCAGGTCTTCTATTTCAAGC 919

Qy        938 TGAC 941

Db        920 TGAC 923


Search completed: May  9, 2004, 15:47:38
Job time : 111 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 12:40:19 ; Search time 492 Seconds
(without alignments)
9351.217 Million cell updates/sec

Title: US-10-089-787-1
Perfect score: 1083
Sequence: 1 atgctcctgctgctggccga.....ccaccttgaagctgcgttga 1083

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1083	100.0	1083	4	AAF86639	Aaf86639 Pseudomon
2	1078.2	99.6	1083	4	AAS54276	Aas54276 Pseudomon
3	1078.2	99.6	1083	7	ACA42599	Aca42599 Prokaryot
4	819.8	75.7	1080	7	ACA43847	Aca43847 Prokaryot
5	767	70.8	1080	7	ACA45659	Aca45659 Prokaryot
6	540.8	49.9	1080	7	ACA31787	Aca31787 Prokaryot
7	539	49.8	1083	4	AAS53957	Aas53957 Klebsiell
8	537.6	49.6	1080	7	ACA35532	Aca35532 Prokaryot
9	518.2	47.8	1083	7	ACA51132	Aca51132 Prokaryot
10	518.2	47.8	1083	7	ACA51964	Aca51964 Prokaryot
11	517.2	47.8	1083	4	AAS56340	Aas56340 Salmonell
12	502.2	46.4	1083	4	AAS52294	Aas52294 E. coli D
13	502.2	46.4	1083	5	AAH81459	Aah81459 Escherich
14	502.2	46.4	1083	7	ACA32358	Aca32358 Prokaryot
15	490.6	45.3	1060	7	ACA49179	Aca49179 Prokaryot
16	480.4	44.4	1167	7	ACA27323	Aca27323 Prokaryot
17	470.2	43.4	1083	7	ACA53405	Aca53405 Prokaryot
18	460.6	42.5	1083	7	ACA53866	Aca53866 Prokaryot
19	440.4	40.7	1167	7	ACA25959	Aca25959 Prokaryot
20	412.6	38.1	1167	7	ACA25051	Aca25051 Prokaryot
21	410	37.9	1083	7	ACA42010	Aca42010 Prokaryot
22	410	37.9	1131	3	AAZ53168	Aaz53168 Neisseria
23	409.2	37.8	1080	7	ABZ41538	Abz41538 N. gonorr
24	409.2	37.8	1080	7	ACA41076	Aca41076 Prokaryot
25	406.2	37.5	1083	7	ACF71681	Acf71681 Photorhab
26	406.2	37.5	110000	7	ACF67367_48	Continuation (49 o
27	406.2	37.5	110000	7	ACF65387_0	Acf65387 Photorhab
c 28	405.2	37.4	69936	3	AAA81479	Aaa81479 N. mening
29	405.2	37.4	110000	3	AAA81490_04	Continuation (5 of
30	405.2	37.4	349980	3	AAF21607	Aaf21607 Neisseria
31	402	37.1	1131	3	AAZ53166	Aaz53166 Neisseria
32	388.6	35.9	1083	7	ACA42842	Aca42842 Prokaryot
33	368	34.0	1080	7	ACA44760	Aca44760 Prokaryot
34	361.4	33.4	1083	4	AAS53452	Aas53452 Haemophil
35	361.4	33.4	1083	7	ACA34365	Aca34365 Prokaryot
36	361.4	33.4	110000	2	AAT42063_11	Continuation (12 o
37	361.4	33.4	110000	2	AAT42063_12	Continuation (13 o
38	359	33.1	4942	5	AAS89816	Aas89816 DNA encod
39	359	33.1	4944	5	AAS89273	Aas89273 DNA encod
40	359	33.1	4944	5	AAS90181	Aas90181 DNA encod
c 41	326.6	30.2	1310	5	AAS93847	Aas93847 DNA encod
42	326	30.1	645	7	ACA24034	Aca24034 Prokaryot
43	310	28.6	1257	5	AAS90011	Aas90011 DNA encod
44	294.6	27.2	433	7	ABZ56666	Abz56666 Aspergill
45	279.8	25.8	1131	3	AAZ53167	Aaz53167 Neisseria

ALIGNMENTS

RESULT 1
AAF86639
ID AAF86639 standard; DNA; 1083 BP.
XX
AC AAF86639;
XX
DT 17-JUL-2001 (first entry)
XX
DE Pseudomonas aeruginosa mraY gene.
XX
KW MraY; phospho-N-acetylmuramoyl-pentapeptide translocase;
KW peptidoglycan biosynthesis; bacterial cell wall; infection;
KW drug screening; antibacterial; Pseudomonad; G+C rich bacterium; ds.
XX
OS Pseudomonas aeruginosa.
XX
FH Key Location/Qualifiers
FT CDS 1. .1083
FT /*tag= a
FT /product= "MraY protein"
FT /function= "Catalyses the transfer of N-acetylmuramic
FT acid peptide to a bactoprenol phosphate carrier molecule
FT in peptidoglycan biosynthesis"
XX
PN WO200125251-A1.
XX
PD 12-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US027056.
XX
PR 04-OCT-1999; 99US-0157580P.
XX
PA (MERI) MERCK & CO INC.
XX
PI El-Sherbeini M, Azzolina B;
XX
DR WPI; 2001-308221/32.
DR P-PSDB; AAB73487.
XX
PT New MraY gene and enzyme of Pseudomonas aeruginosa, useful in vitro
PT assays for screening antibacterial compounds that target cell wall
PT biosynthesis, particularly for screening antibiotics against
PT Pseudomonads.
XX
PS Claim 2; Fig 1; 22pp; English.
XX

CC This sequence represents the Pseudomonas aeruginosa mraY gene, which
CC encodes a phospho-N-acetylmuramoyl-pentapeptide translocase designated
CC MraY. The MraY protein is involved in bacterial cell wall biosynthesis.
CC It catalyses the first step of the membrane cycle of peptidoglycan
CC biosynthesis, i.e., the transfer of an N-acetylmuramic acid peptide to a
CC bactoprenol phosphate carrier molecule. The MraY protein is useful in in
CC vitro assays to screen for antibacterial compounds that target cell wall
CC biosynthesis. Inhibitors of the MraY protein are useful in preventing the
CC growth of Pseudomonads and other G+C rich bacteria. Pseudomonas
CC aeruginosa is an opportunistic pathogen which causes infections in
CC patients with burns, neutropenia, or cystic fibrosis. Primers or probes
CC derived from the mraY gene are useful in nucleic acid amplification-based
CC assays for detecting the presence of a polynucleotide encoding
CC Pseudomonas aeruginosa MraY protein
XX
SQ Sequence 1083 BP; 164 A; 350 C; 333 G; 236 T; 0 U; 0 Other;

Query Match 100.0%; Score 1083; DB 4; Length 1083;
Best Local Similarity 100.0%; Pred. No. 3.5e-209;
Matches 1083; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080

Qy       1081 TGA 1083
              |||
Db       1081 TGA 1083
```

RESULT 2
AAS54276
ID AAS54276 standard; DNA; 1083 BP.
XX
AC AAS54276;
XX
DT 13-FEB-2002 (first entry)
XX
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #407.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU36417.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 7913; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also

```
CC   useful for the identification of potential new targets for antibiotic
CC   development. The antisense nucleic acids can also be used to identify
CC   proteins used in proliferation, to express these proteins, and to obtain
CC   antibodies capable of binding to the expressed proteins. The proteins can
CC   be used to screen compounds in rational drug discovery programmes. The
CC   antisense nucleic acid sequence is also useful to screen for homologous
CC   nucleic acids which are required for cell proliferation in a wide variety
CC   of organisms. The present sequence encodes an essential prokaryotic
CC   cellular proliferation protein. Note: The sequence data for this patent
CC   did not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1083 BP; 163 A; 348 C; 336 G; 236 T; 0 U; 0 Other;

  Query Match           99.6%;  Score 1078.2;  DB 4;  Length 1083;
  Best Local Similarity   99.7%;  Pred. No. 3.3e-208;
  Matches 1080;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATCGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080

Qy       1081 TGA 1083
              |||
Db       1081 TGA 1083


RESULT 3
ACA42599
ID   ACA42599 standard; DNA; 1083 BP.
XX
AC   ACA42599;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #24256.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU38729.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 30469; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
```

CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1083 BP; 163 A; 348 C; 336 G; 236 T; 0 U; 0 Other;

Query Match 99.6%; Score 1078.2; DB 7; Length 1083;
Best Local Similarity 99.7%; Pred. No. 3.3e-208;
Matches 1080; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATCGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080

Qy       1081 TGA 1083
              |||
Db       1081 TGA 1083
```

RESULT 4
ACA43847
ID ACA43847 standard; DNA; 1080 BP.
XX
AC ACA43847;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #25504.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas putida.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU39977.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 31717; 1766pp; English.

XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1080 BP; 169 A; 329 C; 334 G; 248 T; 0 U; 0 Other;

Query Match 75.7%; Score 819.8; DB 7; Length 1080;
Best Local Similarity 85.0%; Pred. No. 4.3e-156;
Matches 917; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||| |||||| ||||||| || ||||||||||| |||| |||||||  ||||| |||
Db          1 ATGCTGCTGCTGTTGGCCGAGTATCTGCAACAGTTCCACAAAGGCTTCGCGGTCTTTCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              |||||| |||||||||| |||||  | || || |||||| |||| ||| ||||||||||||
Db         61 TACCTGTCCCTGCGCGGGATTCTTGGTGTACTGACCGCGTTGTCCCTGGCGCTGTGGCTG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || |||||||||||||||||| |||| || | ||||||||||||||||||| ||||||
Db        121 GGCCCCTGGATGATCCGTACCCTGCAAATTCGCCAGATCGGCCAGGCCGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||||| |||||||||||| |||    |||||||||||||||||| || || ||||||||
Db        181 GGCCCGCAATCGCACCTGTCCAAGTCCGGCACCCCGACCATGGGTGGGGCGCTGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              | ||||| ||| |||||||  |||||||||| || ||   |||||||| || |||||
Db        241 TCGGCCATCGCCGTCAGCACCTTGCTGTGGGCCGACCTGAGCAACCGCTATGTATGGGTT 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||| | || ||||||  |||||||||||| |||||||| || ||||||||||||||||
Db        301 GTGCTGATTGTCACCCTGGCGTTCGGTGCCATTGGCTGGGTCGATGACTACCGCAAGGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||| |||||||| |||||||| |||||||||||||||||||||||| |||||||||
Db        361 ATCGAAAAGAACTCGCGTGGCCTTCCGAGCCGCTGGAAGTACTTCTGGCAATCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||| | |  || || |||||||| |||| ||| ||   ||||    |||||||||| |||
Db        421 GGCCTGGCGGCGGCGGTGTTCCTGTACAAGACGGCGCCAACCAGCGTCGAGACCACGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||| ||||| |  | |||   |||   || | ||||||| ||  |||||| || ||||||
Db        481 ATCCTGCCGTTCATCAAGGATGTCACCATTCCGTTGGGCGTCGGCTTCGTCGTACTGACC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||||||| |||||||||||||| ||||| || |||||||  || || |||||||||
Db        541 TACTTCGTCATTGTCGGCTCGAGCAACGCGGTCAATCTCACCGATGGCCTTGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              || |||||||||||||| ||||| | ||||||||||||||||||||||||||||||||||
Db        601 GCAATCATGCCGACGGTGATGGTCGGCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGT 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              |||||||||||||||| |||||||||| ||| |||| |||||  | |||| ||||||
Db        661 AACGTGAAGTTCGCCGAATACCTGCTGATCCCCTACGTGCCGGGCTCGGGCGAACTGATC 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |||||||||| ||||||| |||| |||||||| ||||| || |||||||||||| ||||||
Db        721 GTGTTCTGCGGCGCGCTGATCGGTGCCGGCCTGGGCTTTCTGTGGTTCAACACTTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||| || |||||||||||||||||||||||||| |||||||||||||||||||||||||
Db        781 GCGCAAGTGTTCATGGGCGACGTCGGCGCGCTGGCACTGGGCGCCGCGCTGGGCACCATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || || |||||||||||||| |||||||||||||||||||| ||  | |||||  |||||
Db        841 GCCGTCATCGTGCGCCAGGAAATCGTGCTGTTCATCATGGGCGGCATCTTCGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||| |||||| |||||||||||| |||||||||||||||||    ||||| |||||
Db        901 ACCCTGTCGGTGGTGATCCAGGTCGCCTCCTTCAAGCTGACCGGCAAGCGCGTGTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||| |||||||| || |||||| || |||||||| || ||  | ||||||||
Db        961 ATGGCGCCGATTCATCACCACTTTGAACTCAAGGGCTGGCCAGAGCCCAGGGTGATCGTC 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCG 1079
              || |||||||||||||||||||||||||||| ||||| ||||| || ||| |||| ||| |
Db       1021 CGTTTCTGGATCATCACCGTGATCCTGGTACTGATTGGCCTTGCAACCCTGAAACTGAG 1079
```

RESULT 5
ACA45659
ID ACA45659 standard; DNA; 1080 BP.
XX
AC ACA45659;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #27316.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Pseudomonas syringae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU41789.

XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 33529; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1080 BP; 168 A; 302 C; 330 G; 280 T; 0 U; 0 Other;

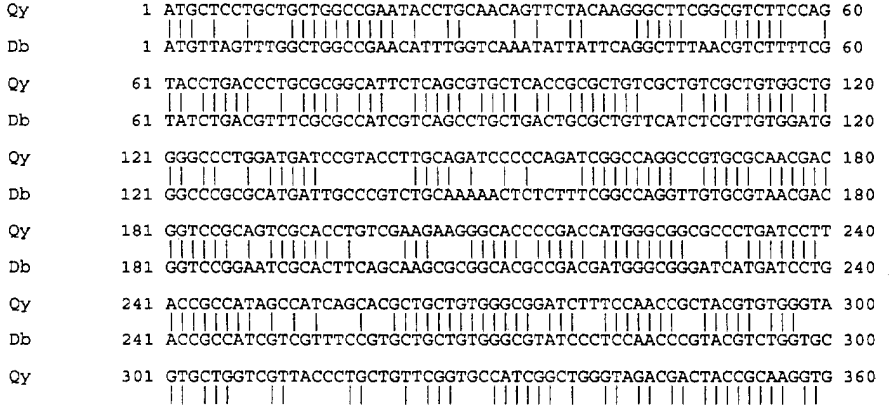
Query Match 70.8%; Score 767; DB 7; Length 1080;
Best Local Similarity 81.9%; Pred. No. 1.9e-145;
Matches 884; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

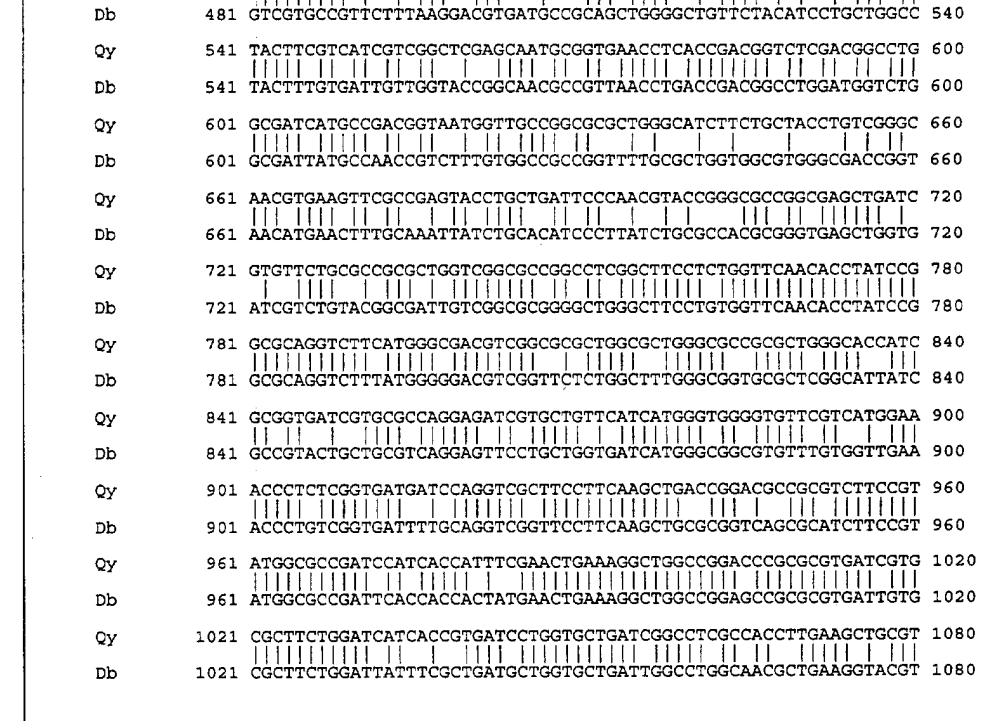
```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||| |||||| ||||||| | | ||||||||||||||| ||||| |  |||||  |||
Db          1 ATGCTGCTGCTGTTGGCCGAGTTCTTACAACAGTTCTACAAAGGCTTTGCGGTCTTTCAG 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              |||||| | |||||||| || ||  |||||||||| |||||| || ||||||||||  |||
Db         61 TACCTGTCTCTGCGCGGGATCCTTGGCGTGCTCACTGCGCTGACGTTGTCGCTGTGCCTG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              |||||  |||||||||||  ||| |||||||  | ||||||||||||| |  ||  ||  |||||
Db        121 GGGCCGTGGATGATCCGCACCCTGCAGATGCGCCAGATCGGCCAGTCGGTCCGTAACGAT 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||||| ||||||  ||||  |||    || || ||||||||||||||  || ||||||||
Db        181 GGTCCGCAATCGCACTTGTCCAAGTCCGGTACGCCGACCATGGGCGGTGCGCTGATCCTC 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              |  ||||  |  ||||||||  |||| |||||  ||| |    ||||||  |||||  |||||
Db        241 TCGTCCATCGGTATCAGCACCTTGCTCTGGGCTGATTTGAGCAACCGTTACGTCTGGGTT 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||  ||  |  ||  |||  | ||||||||||||||||||||||||  || ||||||||||||||
Db        301 GTGTTGCTGGTGACCTTTCTGTTCGGTGCCATCGGCTGGGTTGATGACTACCGCAAGGTC 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||  ||  |||||  |||||||| ||||||||  |||||||||||  ||||||||||||||||
Db        361 ATTGAAAAGAATTCCCGTGGTCTGCCGAGTCGCTGGAAGTATTTCTGGCAGTCGGTGTTC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |||  |  ||||  ||  |  ||  || ||||  ||||||    |  |  |||  |||  |||||   ||
Db        421 GGCCTTTGCGCGGCAATCTTTCTGTACACGACTGCGCCATCGGCGACCGAAACCACATTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||  |||   ||   ||  |  |  |  ||||||   ||||  |  ||  |||||
Db        481 ATCGTGCCGATGCTCAAGGATGTGCGCATTCCGCTCGGCATCGGCTTCATCGTGCTGACG 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||  ||  ||  |||||  ||  |||||  ||  ||  |||||  ||  |||||  ||  |||||  |||
Db        541 TACTTTGTGATTGTCGGTTCCAGCAACGCAGTCAACCTGACTGACGGCCTGGACGGGCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||  |||||  |  |||  ||||||||||||||||||||||||||||||||
Db        601 GCGATCATGCCGACGGTGATGGTCGGCGGTGCGCTGGGCATCTTCTGCTACCTGTCGGGT 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              |||||||  |||||  ||  ||||||||||||||||  ||||  |||||  ||  ||  ||||||||
Db        661 AACGTGAAATTCGCGGAATACCTGCTGATTCCCTACGTGCCGGGGGCGGGTGAGCTGATT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||  ||||  |  ||  |||  |  ||  ||  |||||  ||  ||||||||||||||||||||
Db        721 GTCTTCTCGGGTGCACTGATTGGTGCGGGGCTCGGGTTTCTCTGGTTCAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||  |||||  ||||||||||||||||  ||||||||  ||||||||  ||||||||||||||
Db        781 GCTCAGGTGTTCATGGGCGACGTCGGTGCGCTGGCACTGGGCGCTGCGCTGGGCACCATG 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||  ||||||||||||  |||||  ||  ||  ||||||||||||||  ||  ||||||||  ||||||
Db        841 GCCGTGATCGTGCGTCAGGAAATGGTCCTGTTCATCATGGGCGGCGTGTTCGTGATGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              |||||  ||  |||  |  ||  |||||  ||  ||  |||||  ||||||||  ||||||||  |||||
Db        901 ACCCTTTCAGTGGTCATTCAGGTTGCGTCTTTCAAACTGACCGGCCGCCGCGTATTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              ||||||||||  ||  |||||  ||  |||||  ||||||||||  ||  ||  ||||||||||||
Db        961 ATGGCGCCGATTCACCACCACTTTGAACTCAAAGGCTGGCCCGAGCCACGCGTGATCGTC 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCG 1079
              ||  ||||||||||||||||||  ||||||||  ||||||||  |||||  ||||  |||  |
Db       1021 CGTTTCTGGATCATCACCGTGATTCTGGTGCTCATCGGCCTTGCCACACTGAAACTGAG 1079
```

RESULT 6
ACA31787
ID ACA31787 standard; DNA; 1080 BP.
XX
AC ACA31787;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #13444.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Enterobacter cloacae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.

XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU27917.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 19657; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1080 BP; 169 A; 283 C; 334 G; 294 T; 0 U; 0 Other;

Query Match 49.9%; Score 540.8; DB 7; Length 1080;
Best Local Similarity 68.8%; Pred. No. 7.9e-100;
Matches 743; Conservative 0; Mismatches 337; Indels 0; Gaps 0;





RESULT 7
AAS53957
ID AAS53957 standard; DNA; 1083 BP.
XX
AC AAS53957;
XX
DT 13-FEB-2002 (first entry)
XX
DE Klebsiella pneumoniae DNA for cellular proliferation protein #86.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Klebsiella pneumoniae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU36098.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 7594; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1083 BP; 178 A; 304 C; 323 G; 278 T; 0 U; 0 Other;

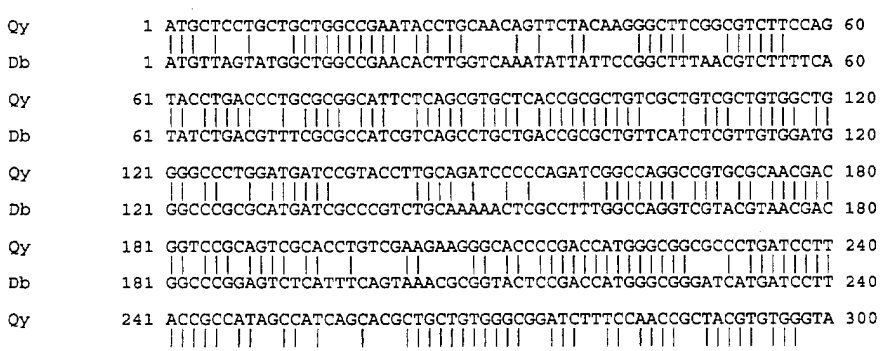
Query Match 49.8%; Score 539; DB 4; Length 1083;
Best Local Similarity 68.6%; Pred. No. 1.8e-99;
Matches 743; Conservative 0; Mismatches 340; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |   |||||||||| || ||    | |  ||    |||||   ||||||
Db          1 ATGTTAGTATGGCTGGCCGAACACTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCA 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| |||| ||||||||||    | ||| |||||  ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCGTTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | ||||||       |||| |  | |    | ||||||| ||| || ||||||
Db        121 GGCCCGCGCATGATCGCCCGTCTGCAAAAACTCGCCTTTGGCCAGGTCGTACGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| |||| ||   |     ||    || || ||||||||||||||    | |||||||
Db        181 GGCCCGGAGTCTCATTTCAGTAAACGCGGTACTCCGACCATGGGCGGGATCATGATCCTC 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              |||||  ||  ||  |   |||||||||||  |||  || ||||  |||||  |||
Db        241 ACCGCGATCACCGTTTCCGTTCTGCTGTGGGCCTATCCATCTAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              || |||   ||      |   | |||   ||||||  |  || || || ||||| || ||
Db        301 GTACTGACGGTATTAATCGGCTACGGCATCATCGGTTTCGTTGATGATTACCGTAAAGTC 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
                 ||  |  ||  ||||||     |||||||||||| ||||||  ||||||| ||
Db        361 GTGCGCAAAGATACCAAAGGCCTGATCGCCCGCTGGAAGTATTTCTGGATGTCGGTGATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              | | | ||||  |||  |    ||  ||  ||  | |   ||     |    |  |||   |||
Db        421 GCCCTGGGCGTGGCCTTCGCGCTGTATCTGGCGGGGAAAGATACCCCGGCAACCGAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||||  |  | ||  |||  |  |||  ||||  | ||||  |  | ||| ||
Db        481 GTGGTGCCGTTCTTTAAAGACGTAATGCCGCAGCTGGGGCTGTTCTATATCTTGCTGGCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||||| ||||||||  |  |||| || || ||||| ||||||| ||||||||||||
Db        541 TACTTCGTTATCGTCGGTACCGGCAACGCCGTCAACCTGACCGACGGCCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              |||||  |||||||| ||  | || || | |  | |  |  |  |  |  | |||
Db        601 GCGATTATGCCGACCGTTTTCGTCGCGGCAGGCTTTGCGCTGGTGGCATGGGCCACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| ||||||  | |||||||  || ||  |  ||     ||| ||||||||| |
Db        661 AACATGAATTTCGCCAACTACCTGCATATCCCGTATCTGCGCCACGCTGGCGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  | ||| | ||||| || || || |||||||| |||||||||||||| |||
Db        721 ATCGTCTGTACGGCGATTGTCGGGGCGGGGCTGGGCTTCCTGTGGTTCAACACCTACCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||| ||||| ||||||||||||  | |||||||| |||| |||| ||||||  |||
Db        781 GCGCAAGTCTTTATGGGCGACGTCGGTTCACTGGCGCTCGGCGGCGCGTTGGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| |||||| || ||||| |||||||| |||| ||||  ||||
Db        841 GCCGTGCTGCTGCGTCAGGAGTTCCTGCTGGTGATCATGGGCGGGGTTTTCGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||||| ||  | |||||||  |||| ||||||  ||| |  ||| |||||||
Db        901 ACGCTGTCGGTTATTCTGCAGGTCGGCTCCTTTAAGCTGCGCGGTCAGCGCATCTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||||||| ||||| |  |||||||||||||||||||| |||||||| ||||||
Db        961 ATGGCGCCGATCCACCACCACTATGAACTGAAAGGCTGGCCGGAGCCGCGCGTTATCGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| ||  |  |||| |||||||||||| ||||| || ||  |||||  | |||
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCGACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
              | |
Db       1081 TAA 1083
```

RESULT 8
ACA35532
ID ACA35532 standard; DNA; 1080 BP.
XX
AC ACA35532;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #17189.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Klebsiella pneumoniae.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU31662.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 23402; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1080 BP; 177 A; 302 C; 322 G; 279 T; 0 U; 0 Other;

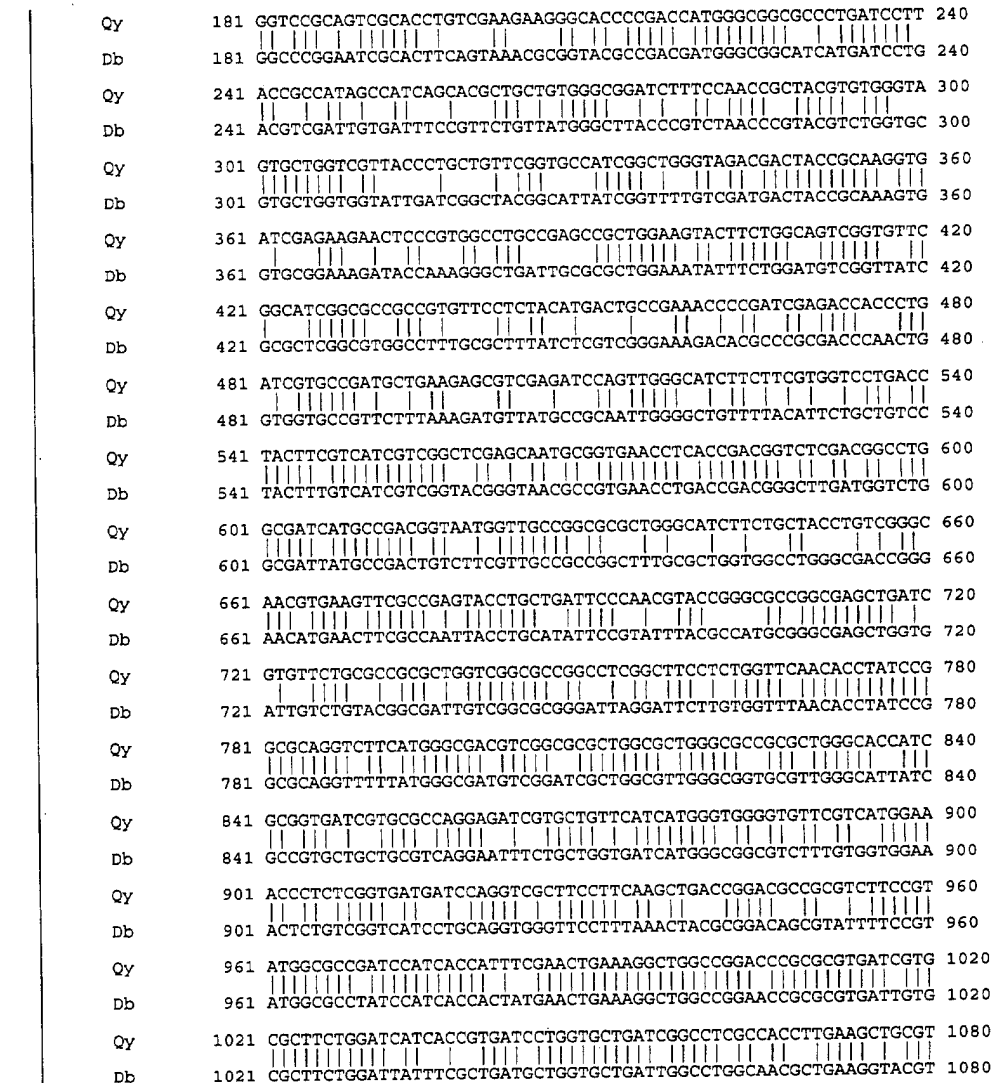
Query Match 49.6%; Score 537.6; DB 7; Length 1080;
Best Local Similarity 68.6%; Pred. No. 3.5e-99;
Matches 741; Conservative 0; Mismatches 339; Indels 0; Gaps 0;



RESULT 9
ACA51132
ID ACA51132 standard; DNA; 1083 BP.
XX
AC ACA51132;
XX
DT 19-JUN-2003 (first entry)
XX
DE Prokaryotic essential gene #32789.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Salmomella typhimurium.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.

XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU47262.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 39002; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1083 BP; 182 A; 262 C; 322 G; 317 T; 0 U; 0 Other;

Query Match 47.8%; Score 518.2; DB 7; Length 1083;
Best Local Similarity 67.4%; Pred. No. 2.9e-95;
Matches 730; Conservative 0; Mismatches 353; Indels 0; Gaps 0;



RESULT 10
ACA51964
ID ACA51964 standard; DNA; 1083 BP.
XX
AC ACA51964;
XX
DT 19-JUN-2003 (first entry)
XX

DE Prokaryotic essential gene #33621.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Salmomella typhi.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU48094.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 39834; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1083 BP; 183 A; 262 C; 322 G; 316 T; 0 U; 0 Other;

Query Match 47.8%; Score 518.2; DB 7; Length 1083;
Best Local Similarity 67.4%; Pred. No. 2.9e-95;
Matches 730; Conservative 0; Mismatches 353; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
Db          1 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db        121 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db        181 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db        241 ACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db        301 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db        361 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db        421 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db        481 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db        541 TACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACCGACGGGCTTGATGGTCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db        601 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCTTGGGCGACCGGG 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db        661 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db        721 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db        781 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGCGCGTTGGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db        841 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db        901 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db        961 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
Db       1081 TAA 1083
```

RESULT 11
AAS56340
ID AAS56340 standard; DNA; 1083 BP.
XX
AC AAS56340;
XX
DT 13-FEB-2002 (first entry)
XX
DE Salmonella typhi DNA for cellular proliferation protein #373.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU38481.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 9977; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1083 BP; 183 A; 262 C; 321 G; 316 T; 0 U; 1 Other;

Query Match 47.8%; Score 517.2; DB 4; Length 1083;
Best Local Similarity 67.3%; Pred. No. 4.6e-95;
Matches 729; Conservative 0; Mismatches 354; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| | | |||||||| | || | | ||  |||||  ||||||
Db          1 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| ||| ||||| || | |||||||||||  | ||  | ||| ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | ||||||  |    |||| |  | |    | |||||||  || || |||||
Db        121 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| | |||||| |    ||    || || ||||| |||||||||  | |||||||
Db        181 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || ||  |  ||  |   ||| | ||||| | |  || ||||  ||||| |||
Db        241 ACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||| ||      |  | |||    |||| |  || || ||||||||||| |||
Db        301 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |  |||  |  ||  || |||       |||||||| || ||||||  ||||||  ||
Db        361 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  ||||||  ||| |   || ||  |    |  ||  | ||  ||  |||| |||
Db        421 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | |||||| |  | ||   ||  |  || |||||  | || | |  | ||| ||
Db        481 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||| |||||||||||  || | || || || ||||| |||||||| || || ||| |||
Db        541 TACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACCGACGGGCTTGATGGTCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||| |||||||| ||  | |||||||  ||  |  |  |  |  | | ||
Db        601 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCTTGGGCGACCGGG 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| |||||| | |||||||  |||||  |  |||     || ||||||||| |
Db        661 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              | ||||  | ||| | |||||||| ||  |  || ||| | |||||  ||||||||||||
Db        721 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||| || |||||||| |||||  |||||||  |||||  |||| ||||||  |||
Db        781 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGCGCGTTGGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| |||||  |  |||||  | |||||||| || || || ||  |||||
Db        841 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||||| ||  | ||||| | |||||| || ||   |||||  ||  | ||||||
Db        901 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||| |||||||||| |  |||||||||||||||||| ||||||||||| |||
Db        961 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| ||  | |||| |||||||||| ||||| || ||  ||||  | |||
Db       1021 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAANGTACGT 1080

Qy       1081 TGA 1083
              | |
Db       1081 TAA 1083
```

RESULT 12
AAS52294
ID AAS52294 standard; DNA; 1083 BP.
XX
AC AAS52294;
XX
DT 13-FEB-2002 (first entry)
XX
DE E. coli DNA for cellular proliferation protein #16.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
KW antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU34435.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Claim 27; SEQ ID NO 5931; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence encodes an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1083 BP; 190 A; 260 C; 318 G; 315 T; 0 U; 0 Other;

Query Match 46.4%; Score 502.2; DB 4; Length 1083;
Best Local Similarity 66.5%; Pred. No. 4.8e-92;
Matches 720; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  ||||||||||| |  ||   | |  ||   |||||  ||||||
Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
               || |||||  | |||| |||  ||||| |||| ||||||||||   | ||  ||||| ||

Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | |||||   |  |||||  |  |      | || ||||  |||||  ||||||

Db        121 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | ||  |||  |    |||   ||  ||  ||||||||||||||||      |||||||

Db        181 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || ||  |  |||  |  ||||||||||  |  |  |||||  |  |||||  |||

Db        241 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||  ||||  ||      |  |  |||||  ||  ||||  ||  ||  ||  ||||| |||

Db        301 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
                |    ||  ||  ||   ||  ||       ||  |||||||  ||||||  ||||||  |

Db        361 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  | || | ||||  |  ||| ||| |  | | | ||  |  ||  ||  ||     |||

Db        421 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | || ||  |  |  ||   ||  |   ||| ||||  | ||||  |  |  | ||| |

Db        481 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||| || ||  |  |||| |||||  |||||  ||||||||||||||||

Db        541 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              || ||  |||||||| ||| | || |||||  |  |  |  |  |  |  | | |||

Db        601 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||  |||| || |||   ||| |||  || ||  |  | |   ||||||  || ||| |

Db        661 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  ||||| | ||||| || || || |||||||| ||||| ||||||||||||

Db        721 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||||| |||||||| || ||  |||||||  | ||  |  |||  | ||||  |||

Db        781 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||  |  | || |||||  || |||||  | ||  |||||  || ||||||||  | |||

Db        841 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||  ||  ||  |  |||||||  |||||  ||  |||  |||||  ||  | |||||

Db        901 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||  |||||  ||||||||  |  ||||||||||||||||||  ||||||||  ||  |||

Db        961 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              || ||||||||  ||  |  ||||  |||||  |||||  ||  ||  ||  ||  ||||  | |||

Db       1021 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
                |||

Db       1081 TAA 1083
```

RESULT 13
AAH81459

```
ID   AAH81459 standard; DNA; 1083 BP.
XX
AC   AAH81459;
XX
DT   21-SEP-2001  (first entry)
XX
DE   Escherichia coli protein encoding nucleotide sequence SEQ ID NO:258.
XX
KW   Escherichia coli; identification; proliferation; microorganism;
KW   antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW   bacterial growth inhibition; ds.
XX
OS   Escherichia coli.
XX
PN   WO200148209-A2.
XX
PD   05-JUL-2001.
XX
PF   19-DEC-2000; 2000WO-US034419.
XX
PR   23-DEC-1999;   99US-0173005P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen KL,  Zyskind JW;
XX
DR   WPI; 2001-457376/49.
DR   P-PSDB; AAG98403.
XX
PT   Novel nucleic acids encoding proteins required for Escherichia coli
PT   proliferation, useful for screening for antimicrobial agents.
XX
PS   Claim 9; Page 386-387; 596pp; English.
XX
CC   The present invention describes a purified or isolated nucleic acid
CC   sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC   given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC   microorganism is capable of inhibiting proliferation of a microorganism.
CC   (I) have antibacterial and antibiotic activities, and can be used in gene
CC   therapy. Expression of (I) in a microorganism inhibits proliferation of
CC   the microorganism, and the manufactured antibiotic is useful for reducing
CC   the activity or level of a gene product required for proliferation of a
CC   microorganism in a subject, specifically humans. The nucleic acids that
CC   inhibit bacterial growth or proliferation can be used as antisense
CC   therapeutics for killing bacteria. In addition to therapeutic
CC   applications, the nucleic acid sequences complementary to sequences
CC   required for proliferation can be used as diagnostic tools. For example,
CC   nucleic acid probes complementary to proliferation-required sequences
CC   that are specific for particular species of microorganisms can be used as
CC   probes to identify particular microorganism species in clinical
CC   specimens. AAH81295 to AAH81487 encode the Escherichia coli proteins
CC   given in AAG98239 to AAG98431, and AAH81488 to AAH81491 represent
CC   oligonucleotides, which are used in the exemplification of the present
CC   invention
XX
SQ   Sequence 1083 BP; 190 A; 260 C; 318 G; 315 T; 0 U; 0 Other;

  Query Match             46.4%;  Score 502.2;  DB 5;  Length 1083;
  Best Local Similarity   66.5%;  Pred. No. 4.8e-92;
  Matches  720;  Conservative    0;  Mismatches  363;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |   |||||||||| |  ||    | |  ||    |||||   ||||||
Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||  |||||  | ||||  |||  |||||  ||||  ||||||||||    |  ||  |||||  ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||  ||  | |||||    |   |||||  |  |       | ||  ||||  |||||  ||||||
Db        121 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | ||  |||  |    |||    ||  ||  ||||||||||||||     |||||||
Db        181 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||  ||  ||  |  |||  |   ||||||||||  |  |  ||||  |  |||||  |||
Db        241 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||  ||||  ||      |    | |||||  ||  ||||  ||  ||  ||  ||  |||||  |||
Db        301 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |    ||  ||  ||    ||  ||        ||  ||||||||  ||||||  ||||||  |
Db        361 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |    | ||  |  ||||  |    |||  |||  |  |  |  ||  |  ||  ||  ||     |||
Db        421 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              |  ||  ||    |  |  ||    ||    |    |||  ||||   |  ||||  |  |  |  |||  |
Db        481 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||| ||  ||  |  ||||  |||||  |||||  |||||  ||||||||||||||
Db        541 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              || ||  |||||||| |||  |  ||  |||||  |    |  |    |  |    |    |  |||
Db        601 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||  ||||  ||  |||    |||  |||  ||  ||    |  |     ||||||  ||  |||  |
Db        661 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  |||||  |  |||||  ||  ||  ||  ||||||||  |||||  ||||||||||||
Db        721 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||  ||||||||  ||  ||  ||||||||  |  ||  |  |||  |  ||||  |||
Db        781 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||  ||  |  | ||  |||||  ||  |||||  |  ||  |||||  ||  ||||||||  |  |||
Db        841 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||  ||  ||  ||  ||  |  |||||||  |||||  ||  |||  |||||  ||  |  |||||
Db        901 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||  |||||  ||||||||  |  ||||||||||||||||||||  ||||||||  ||  |||
Db        961 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||  ||||||||  ||  |  ||||  |||||  |||||  ||  ||  ||  ||  |||||  |  |||
Db       1021 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
              | |
Db       1081 TAA 1083


RESULT 14
ACA32358
ID   ACA32358 standard; DNA; 1083 BP.
XX
AC   ACA32358;
XX
DT   19-JUN-2003  (first entry)
XX
```

DE Prokaryotic essential gene #14015.
XX
KW Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Escherichia coli.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU28488.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 14; SEQ ID NO 20228; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1083 BP; 190 A; 260 C; 318 G; 315 T; 0 U; 0 Other;

Query Match 46.4%; Score 502.2; DB 7; Length 1083;
Best Local Similarity 66.5%; Pred. No. 4.8e-92;
Matches 720; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||| | ||   | |  ||   |||||  ||||||

Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| |||| ||||||||||   | ||  ||||| ||

Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTCGCGCAACGAC 180
              || ||  | |||||   |   |||||  |  |      | || ||||  ||||| ||||||

Db        121 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | || |||  |    |||   || || ||||||||||||||     |||||||

Db        181 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || ||  |  |||  |   |||||||||||  | |  |||||  |  ||||| |||

Db        241 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||| |||| ||     |   | |||||  || ||||  || || || || ||||| |||

Db        301 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |   ||  || ||  ||  ||      || |||||||| ||||||  ||||||  |

Db        361 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |   | || | |||| |   ||| ||| |  | | | ||  | ||  ||  ||    |||

Db        421 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | || ||  |  | ||   ||   |  ||| |||| | |||| | |  | | ||| |

Db        481 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||||||| || ||  | |||| ||||| ||||| ||||| ||||||||||||||||

Db        541 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              || || |||||||| ||| | || ||||| |  | |  |  |  |  |  | | ||||

Db        601 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||  |||| || |||   ||| |||  || ||  |  | |   ||||| || ||| |

Db        661 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  |||| ||||| | ||||| || || || |||||||| ||||| ||||||||||||

Db        721 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||||| |||||||| || ||  |||||||| | || |  ||| | ||||  |||

Db        781 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||  |  || |||||  || ||||| | || ||||| || ||||||||  | |||

Db        841 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || || || ||  | |||||||  |||||  || |||  |||||  ||  | |||||

Db        901 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              ||||| ||||| |||||||| | ||||||||||||||||||||| ||||||||  || |||

Db        961 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              || |||||||  ||  | |||| ||||| ||||| || || || ||  ||||| | |||

Db       1021 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
              |||
Db       1081 TAA 1083
```

```
RESULT 15
ACA49179
ID   ACA49179 standard; DNA; 1060 BP.
XX
AC   ACA49179;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Prokaryotic essential gene #30836.
XX
KW   Antisense; ds; prokaryotic essential gene; cell proliferation;
KW   drug design; gene.
XX
OS   Salmonella paratyphi.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,  Zamudio C,  Malone C,  Haselbeck R,  Ohlsen KL,  Zyskind JW;
PI   Wall D,  Trawick JD,  Carr GJ,  Yamamoto R,  Forsyth RA,  Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   P-PSDB; ABU45309.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 14; SEQ ID NO 37049; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is one of the target
CC   prokaryotic essential genes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1060 BP; 180 A; 258 C; 313 G; 309 T; 0 U; 0 Other;

  Query Match           45.3%;  Score 490.6;  DB 7;  Length 1060;
  Best Local Similarity   68.2%;  Pred. No. 1.1e-89;
  Matches  710;  Conservative    0;  Mismatches  329;  Indels    2;  Gaps    2;

Qy         43 GGCTTCGGCGTCTTCCAGTACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTG 102
Db         19 GGCTTTAACGTCTTTTCTTATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTG 78

Qy        103 TCGCTGTCGCTGTGGCTGGGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGC 162
Db         79 TTCATCTCTTTATGGATGGGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGC 138

Qy        163 CAGGCCGTGCGCAACGACGGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATG 222
Db        139 CAGGTGGTACGTAACGATGGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATG 198

Qy        223 GGCGGCGCCCTGATCCTTACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCC 282
Db        199 GGCGGCATCATGATCCTGACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCT 258

Qy        283 AACCGCTACGTGTGGGTAGTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTA 342
Db        259 AACCCGTACGTCTGGTGCGTGCTGGTGGTATTGATTGGCTACGGCATTATCGGTTTTGTC 318

Qy        343 GACGACTACCGCAAGGTGATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTAC 402
Db        319 GATGACTACCACAAAGTGGTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATAT 378

Qy        403 TTCTGGCAGTCGGTGTTCGGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACC 462
Db        379 TTCTGGATGTCGGTTATCGCGCTCGG-GTGGCCTTTGCGCTTTATCTCGTCGGGAAAGAC 437

Qy        463 CCGATCGAGACCACCCTGATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATC 522
Db        438 ACGCCCGCGACCCAACTGGTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTG 497

Qy        523 TTCTTCGTGGTCCTGACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACC 582
Db        498 TTTTACATTCTGCTGTCCTACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACC 557

Qy        583 GACGGTCTCGACGGCCTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATC 642
Db        558 GACGGGCTTGATGGTCTGGCGATTATGCCGACTGTTTTCGTTGCCGCCGGCTTTGCGCTG 617

Qy        643 TTCTGCTACCTGTCGGGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCG 702
Db        618 GTGGCCTGGGCGACCGGGAACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGC 677

Qy        703 GGCGCCGGCGAGCTGATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTC 762
Db        678 TATGCGGGCGAGCTGGTGATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTG 737

Qy        763 TGGTTCAACACCTATCCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGC 822
Db        738 TGGTTTAACACCTATCCGGCGCAGGTTTTTATGGGCGACGTCGGATCGCTGGCGTTGGGC 797

Qy        823 GCCGCGCTGGGCACCATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGT 882
Db        798 GGCGCGTTGGGCATTATCGCCGTGCTGCTGCGTCAGGAGTTCCTGCTGGTGATCATGGGC 857

Qy        883 GGGGTGTTCGTCATGGAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACC 942
Db        858 GGCGTCTTTGTGGTGGAAACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGC 917

Qy        943 GGACGCCGCGTCTTCCGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCG 1002
Db        918 GGACAGC-TATTTTCCGTATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCG 976

Qy       1003 GACCCGCGCGTGATCGTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTC 1062
```

```
                || |||||||||| |||||||||||||| ||  |  |||| |||||||||| |||||
Db        977 GAACCGCGCGTGATTGTGCGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTG 1036

Qy       1063 GCCACCTTGAAGCTGCGTTGA 1083
                || ||  ||||| | |||| |
Db       1037 GCAACGCTGAAGGTACGTTAA 1057
```

Search completed: May 9, 2004, 13:38:24
Job time : 495 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 13:17:24 ; Search time 4393 Seconds
(without alignments)
10685.296 Million cell updates/sec

Title: US-10-089-787-1
Perfect score: 1083
Sequence: 1 atgctcctgctgctggccga.....ccaccttgaagctgcgttga 1083

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1083	100.0	5280	1	AY008276	AY008276 Pseudomon
2	c	1078.2	99.6	24000	1	AE004856	AE004856 Pseudomon
3		820.2	75.7	301995	1	AE016779	AE016779 Pseudomon
4	c	767.4	70.9	311600	1	AE016871	AE016871 Pseudomon
5		598	55.2	13045	1	AE012171	AE012171 Xanthomon
6		590	54.5	13079	1	AE011708	AE011708 Xanthomon
7	c	585.8	54.1	208524	1	AE016925	AE016925 Chromobac
8		518.2	47.8	22348	1	AE008699	AE008699 Salmonell
9		518.2	47.8	251050	1	AL627265	AL627265 Salmonell
10		518.2	47.8	300169	1	AE016834	AE016834 Salmonell
11		502.2	46.4	1083	6	AX189057	AX189057 Sequence
12		502.2	46.4	2608	1	ECMUROY	X51584 Escherichia
13		502.2	46.4	21757	1	AE000118	AE000118 Escherich
14		502.2	46.4	28277	6	AX191720	AX191720 Sequence
15		502.2	46.4	28277	15	EC2MIN	X55034 E. coli 2 m
16		502.2	46.4	111408	1	ECO110K	D10483 Escherichia
17		502.2	46.4	300409	1	AE016755	AE016755 Escherich
18		494.2	45.6	12791	1	AE005185	AE005185 Escherich
19		494.2	45.6	281530	1	AP002550	AP002550 Escherich
20		491	45.3	11441	1	AE015046	AE015046 Shigella
21		491	45.3	290029	1	AE016978	AE016978 Shigella
22	c	487.6	45.0	348074	1	BX640449	BX640449 Bordetell
23	c	486	44.9	344805	1	BX640434	BX640434 Bordetell
24	c	482.8	44.6	348134	1	BX640420	BX640420 Bordetell
25	c	473.2	43.7	188050	1	AL646072	AL646072 Ralstonia
26	c	470.2	43.4	22201	1	AE004310	AE004310 Vibrio ch
27		470.2	43.4	249150	1	AP005332	AP005332 Vibrio vu
28		470	43.4	316050	1	BX321859	BX321859 Nitrosomo
29	c	467	43.1	301442	1	AE016798	AE016798 Vibrio vu
30	c	460.6	42.5	11726	1	AE013965	AE013965 Yersinia
31		460.6	42.5	201050	1	AJ414143	AJ414143 Yersinia
32		459	42.4	298900	1	AP005074	AP005074 Vibrio pa
33	c	456.8	42.2	298600	1	AP005959	AP005959 Bradyrhiz
34		455	42.0	1032	6	AR388539	AR388539 Sequence
35		442.4	40.8	208805	2	AC147366	AC147366 Mus muscu
36	c	441.4	40.8	13071	1	AE015855	AE015855 Shewanell
37		438.2	40.5	21000	1	AB052554	AB052554 Shewanell
38	c	425.6	39.3	348068	1	BX572604	BX572604 Rhodopseu
39	c	425.4	39.3	301205	1	AE017218	AE017218 Geobacter
40		420.4	38.8	20062	1	AE003919	AE003919 Xylella f
41	c	420.4	38.8	301443	1	AE012560	AE012560 Xylella f
42	c	418.4	38.6	329709	1	AP002997	AP002997 Mesorhizo
43	c	410	37.9	195767	1	NMA7Z2491	AL162758 Neisseria
44	c	406.2	37.5	349944	1	BX571871	BX571871 Photorhab
45		406.2	37.5	349980	6	AX770909	AX770909 Sequence

ALIGNMENTS

RESULT 1
AY008276
LOCUS AY008276 5280 bp DNA linear BCT 19-JUL-2001
DEFINITION Pseudomonas aeruginosa MurE (murE), MurF (murF), MraY (mraY), and MurD (murD) genes, complete cds.
ACCESSION AY008276
VERSION AY008276.2 GI:12007715
KEYWORDS .
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 5280)
AUTHORS Azzolina,B.A., Yuan,X., Anderson,M.S. and El-Sherbeini,M.
TITLE The cell wall and cell division gene cluster in the Mra operon of

```
            Pseudomonas aeruginosa: cloning, production, and purification of
            active enzymes
  JOURNAL   Protein Expr. Purif. 21 (3), 393-400 (2001)
  MEDLINE   21178826
   PUBMED   11281713
REFERENCE   2  (bases 1 to 5280)
  AUTHORS   El-Sherbeini,M. and Azzolina,B.
  TITLE     Pseudomonas aeruginosa mraY, murD, murE, and murF genes
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 5280)
  AUTHORS   El-Sherbeini,M. and Azzolina,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2000) Biochemistry, Merck Research Laboratories,
            126 E. Lincoln Avenue, Rahway, NJ 07065, USA
REFERENCE   4  (bases 1 to 5280)
  AUTHORS   El-Sherbeini,M. and Azzolina,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2001) Biochemistry, Merck Research Laboratories,
            126 E. Lincoln Avenue, Rahway, NJ 07065, USA
  REMARK    Sequence update by submitter
COMMENT     On Jan 2, 2001 this sequence version replaced gi:10719753.
FEATURES             Location/Qualifiers
     source          1. .5280
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:287"
     gene            1. .1464
                     /gene="murE"
     CDS             1. .1464
                     /gene="murE"
                     /note="UDP-N-acetylmuramoyl-L-alanyl-D-glutamate:meso-2,
                     6-diaminopimelate ligase"
                     /codon_start=1
                     /transl_table=11
                     /product="MurE"
                     /protein_id="AAG45235.1"
                     /db_xref="GI:12007716"
                     /translation="MPMSLSQLFPQAERDLLIRELTLDSHGVRPVDLFLTVPGGHQDG
                     RAHIADALTKGATAVAYEAEGAGELPPSDAPLIAVKGLAAQLSAVAGRFYGEPSRGLD
                     LIGVTGTNGKTSVSQLVAQALDLLGERCGIVGTLGTGFYGALESGRHTTPDPLAVQAT
                     LATLKQAGARAVAMEVSSHGLDQGRVAALGFDIAVFTNLSRDHLDYHGSMEAYAAAKA
                     KLFAWPDLRCRVINLDDDFGRRLAGEEQDSELITYSLTDSSAFLYCREASPTASPTAR
                     RSSIAAKQGLLRSPLLGRFNLSNLLAAVGALLGLGYPLGDILRTLPQLQGPVGRMQRL
                     GGGGKPLVVVDYAHTPDALEKVLEALRPHAAARLLCLFGCGGDRDAGKRPLMAAIAER
                     LADEVLVTDDNPRTEASAAIIADIRKGFAAADKVTFLPSRGEAIAHLIASAAVDDVVL
                     LAGKGHEDYQEIDGVRHPFSDIEQAERALAAWEVPHA"
     gene            1457. .2833
                     /gene="murF"
     CDS             1457. .2833
                     /gene="murF"
                     /note="recombinant enzyme catalyzed the ATP-dependent
                     addition of D-alanine-D-alanine to the
                     UDP-N-acetylmuramyl-L-alanine-D-glutamate-m-Dap precursor"
                     /codon_start=1
                     /transl_table=11
                     /product="MurF"
                     /protein_id="AAG45236.1"
                     /db_xref="GI:12007717"
                     /translation="MLEPLRLSQLTVALDARLIGEDAVFSAVSTDSRAIGPGQLFIAL
                     SGPRFDGHDYLAEVAAKGAVAALVEREVADAPLPQLLVRDTRAALGRLGALNRRKFTG
                     PLAAMTGSSGKTAVKEMLASILRTQAGDAESVLATRGNLNNDLGVPLTLLQLAPQHRS
                     AVIELGASRIGEIAYTVELTRPHVAIITNAGTAHVGEFGGPEKIVEAKGEILEGLAAD
                     GTAVLNLDDKAFDTWKARASGRPLLTFSLDRPQADFRAADLQRDARGCMGFRLQGVAG
                     EAQVQLNLLGRHNVANALAAAAAAAHALGVPLDGIVAGLQALQPVKGRAVAQLTASGLR
                     VIDDSYNANPASMLAAIDILSGFSGRTVLVLGDMGELGSWAEQAHREVGAYAAGKVSA
                     LYAVGPLMAHAVQAFGATGRHFADQASLIGALATEQPTTTILIKGSRSAAMDKVVAAL
                     CGSSEESH"
     gene            2833. .3915
                     /gene="mraY"
     CDS             2833. .3915
                     /gene="mraY"
                     /note="phospho-N-acetylmuramoyl-pentapeptide translocase"
                     /codon_start=1
                     /transl_table=11
                     /product="MraY"
                     /protein_id="AAG22121.1"
                     /db_xref="GI:10719754"
                     /translation="MLLLLAEYLQQFYKGFGVFQYLTLRGILSVLTALSLSLWLGPWM
                     IRTLQIPQIGQAVRNDGPQSHLSKKGTPTMGGALILTAIAISTLLWADLSNRYVWVVL
                     VVTLLFGAIGWVDDYRKVIEKNSRGLPSRWKYFWQSVFGIGAAVFLYMTAETPIETTL
                     IVPMLKSVEIQLGIFFVVLTYFVIVGSSNAVNLTDGLDGLAIMPTVMVAGALGIFCYL
                     SGNVKFAEYLLIPNVPGAGELIVFCAALVGAGLGFLWFNTYPAQVFMGDVGALALGAA
                     LGTIAVIVRQEIVLFIMGGVFVMETLSVMIQVASFKLTGRRVFRMAPIHHHFELKGWP
                     DPRVIVRFWIITVILVLIGLATLKLR"
     gene            3929. .5275
                     /gene="murD"
     CDS             3929. .5275
                     /gene="murD"
                     /note="UDP-N-acetylmuramyl-L-alanine:D-glutamate ligase;
                     recombinant enzyme catalyzed the ATP-dependent
                     addition of D-glutamate to the precursor sugar peptide"
                     /codon_start=1
                     /transl_table=11
                     /product="MurD"
                     /protein_id="AAG45237.1"
                     /db_xref="GI:12007718"
                     /translation="MSLIASDHFRIVVGLGKSGMSLVRYLARRGLPFAVVDTRENPPE
                     LATLRAQYPQVEARCGELDAEFLCSARELYVSPGLSLRTPALVQAAAKGVRISGDIDL
                     FAREAKAPIVAITGSNAKSTVTTLVGEMAVAADKRVAVGGNLGTPALDLLADDIELYV
                     LELSSFQLETCDRLNAEVATVLNVSEDHMDRYDGMADYHLAKHRIFRGARQVVVNRAD
                     ALTRPLIADTVPCWSFGLNKPDFKAFGLIEEDGQKWLAFQFGKLLPVGELKIRGAHNY
                     SNALAALALGHAVGLPFDAMLGALKAFSGLAHRCQWVRERQGVSYYDDSKATNVGAAL
                     AAIEGLGADIDGKLVLLAGGDGKGADFHDLREPVARFCRAVVLLGRDAGLIAQALGNA
                     VPLVRVATLDEAVRQAAELAREGDAVLLSPACASLDMFKNFEERGRLFAKAVEELA"
ORIGIN

  Query Match           100.0%;  Score 1083;  DB 1;  Length 5280;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-131;
  Matches 1083;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2833 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 2892

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2893 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 2952

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2953 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 3012

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3013 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 3072

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3073 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 3132

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3133 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 3192

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3193 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 3252

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3253 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 3312

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3313 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 3372
```

```
Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3373 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 3432

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3433 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 3492

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3493 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 3552

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3553 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 3612

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3613 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 3672

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3673 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 3732

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3733 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 3792

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3793 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 3852

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3853 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 3912

Qy       1081 TGA 1083
              |||
Db       3913 TGA 3915
```

RESULT 2
AE004856/c

```
LOCUS       AE004856               24000 bp    DNA     linear   BCT 19-FEB-2003
DEFINITION  Pseudomonas aeruginosa PAO1, section 417 of 529 of the complete
            genome.
ACCESSION   AE004856 AE004091
VERSION     AE004856.1  GI:9950633
KEYWORDS    .
SOURCE      Pseudomonas aeruginosa PAO1
  ORGANISM  Pseudomonas aeruginosa PAO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 24000)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z. and Paulsen,I.T.
  TITLE     Complete genome sequence of Pseudomonas aeruginosa PA01, an
            opportunistic pathogen
  JOURNAL   Nature 406 (6799), 959-964 (2000)
  MEDLINE   20437337
   PUBMED   10984043
REFERENCE   2  (bases 1 to 24000)
  AUTHORS   Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrener,P.,
            Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
            Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
            Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
            Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
            Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
            Hancock,R.E.W., Lory,S. and Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Department of Medicine and Genetics,
            University of Washington Genome Center, University Of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 24000)
  AUTHORS   .
  CONSRTM   Pseudomonas aeruginosa Community Annotation Project (PseudoCAP)
  TITLE     Direct Submission
  JOURNAL   Submitted (04-FEB-2003) Department of Molecular Biology and
            Biochemistry, Simon Fraser University, 8888 University Dr.,
            Burnaby, British Columbia V5A 1S6, Canada
COMMENT     ----------------------------------------------------------------
            This represents the February 3, 2003 version of the continually
            updated, reviewed, Pseudomonas aeruginosa PAO1 genome annotation,
            from PseudoCAP (see http://www.pseudomonas.com for latest updates
            and links to alternate annotations). PseudoCAP is coordinated by
            Fiona S.L. Brinkman (Simon Fraser University, Canada) and Robert
            E.W. Hancock (University of British Columbia, Canada). We welcome
            submission through www.pseudomonas.com of any proposed changes.

            'Protein name confidence' is used to rate our confidence of the
            accuracy of the protein name.
            Class 1: Function experimentally demonstrated in P. aeruginosa.
            Class 2: Function of highly similar gene experimentally
                 demonstrated in another organism (and gene context consistent
                 in terms of pathways its involved in, if known).
            Class 3: Function proposed based on presence of conserved amino
                 acid motif, structural feature or limited sequence similarity
                 to an experimentally studied gene.
            Class 4: Homologs of previously reported genes of unknown function,
                 or no similarity to any previously reported sequences.
            ----------------------------------------------------------------.
FEATURES             Location/Qualifiers
     source          1. .24000
                     /organism="Pseudomonas aeruginosa PAO1"
                     /mol_type="genomic DNA"
                     /strain="PAO1"
                     /db_xref="taxon:208964"
     gene            complement(91. .1308)
                     /gene="argJ"
                     /locus_tag="PA4402"
     CDS             complement(91. .1308)
                     /gene="argJ"
                     /locus_tag="PA4402"
                     /note="Protein name confidence: Class 2 (high similarity
                     to functionally studied protein); Subcellular
                     localization: Cytoplasmic; Subcellular localization
                     confidence: Class 2; Alt protein name: ornithine
                     transacetylase"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate N-acetyltransferase"
                     /protein_id="AAG07790.1"
                     /db_xref="GI:9950634"
                     /translation="MAVGLGPLSTLHPVPGFELGIASAGIKRPGRKDVVVMRCAEGSS
                     VAGVFTLNAFCAAPVTLAKQRFLGEVRYLLTNTGNANAGTGEAGLAAAAQTCAKLAEL
                     AGVAETSVLPYSTGVIGEPLPVAKIEAALPAALADLAEDRWAEAAAGIMTTDTLPKGA
                     SRQFVHDGVTVTVTGISKGAGMIKPNMATMLGYIATDAKVAQGVLQDLLRDAANKSFN
                     RITIDGDTSTNDCCMLIATGRAALPEVTQASGALFAALKQAVLEVSMELAQAIVRDGE
                     GATKFVTVQVNGGATHQECLDVGYAVAHSPLIKTALFASDPNWGRILAAVGRAGVANL
                     DVSKIDVFLGDVCIASRGGRAASYTEEQGAAVMAQAEIGIRIELGRGTCSETIWTTDL
                     SHEYVKINAEYRT"
     gene            complement(1455. .4205)
                     /gene="secA"
                     /locus_tag="PA4403"
     CDS             complement(1455. .4205)
                     /gene="secA"
                     /locus_tag="PA4403"
                     /note="Protein name confidence: Class 2 (high similarity
                     to functionally studied protein)"
                     /codon_start=1
                     /transl_table=11
                     /product="secretion protein SecA"
```

/protein_id="AAG07791.1"
/db_xref="GI:9950635"
/translation="MFAPLLKKLFGSKNERDVKRMAKAVQAINALEPQMVALSDEQLK
AKTAEFQQRYAKGETLDQLLPEAFAVVREAGKRVMGMRHFDVQLIGGMTLHDGKIAEM
RTGEGKTLVGTLPVYLNALSGKGVHVVTVNDYLARRDANWMRPLYEFLGLSVGVVTPF
QPPEDKRAAYAADITYGTNNEFGFDYLRDNMAFSLDDKFQRELNFAVVDEVDSILIDE
ARTPLIISGQAEDSSELYIKINKLIPRLKRQVEEVEGKPTEEGHYSIDEKTRQVELNE
QGHQFIEDLLSQNGLLGEGESLYSAHNLSLLTHVYAALRAHTLFHRNVEYIVQGDQIL
LIDEHTGRTMPGRRLSEGLHQAIEAKEGLPIQAESQTLASTTFQNYFRLYNKLAGMTG
TADTEAFEFRQIYGLDVVVIPTHRPIARKDFNDLVYLTQEEKYAAIITDIKQCQALGR
PILVGTASIESSEYVSKLLQEAGIEHKVLNAKYHEKEAEIIAQAGAPGSVTIATNMAG
RGTDILLGGNWEVEVAALENPTEEQIAQIKAEWQKRHQQVIEAGGLHVIASERHESRR
IDNQLRGRAGRQGDPGSSRFYLSLEDNLMRIFASDRVKNFMKALGMQSGEAIEHRMVT
NAIEKAQRKVEGRNFDIRKQLLEFDDVANEQRKVIYHMRNTLLSAEDVGETIKEFREE
TLSATINQHIPPQSLPEQWDIEGLEAALYSDFAVRLPIQQWLDEDDKLYEETLRSKIL
EQIVAAYYEKEELAGAEALRAFEKQMLLRVLDDLWKDHLSTMDHLRHGIHLRGYAQKN
PKQEYKRESFTLFQELLDSIKRDTIRVLSHVQVRREDPAEEEARLRREAEELAKRMQF
QHAEAPSMEQAVAGEEEELPEGPAPVVPLEPVRNEQKIGRNEPCPCGSGKKYKHCHGQ
LD"

gene complement(4339. .5262)
/locus_tag="PA4404"

CDS complement(4339. .5262)
/locus_tag="PA4404"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07792.1"
/db_xref="GI:9950636"
/translation="MHIIFLSRHHGAARSVHLQPRLLMLAGACVVGASLLAGGFMGAW
LFRAPAPVAGIEVDAALQAQSAEISSAREDAQRQMDALSVHLGELQARLVRLDALGER
LTELADVDPDEFDFSLSVGQGGVDEPLEGSAYAPPSFMATLDELAARLDTREQQLEVL
EQLLADKHISQSTYLAGQPVRQGYISSPFGRRVNPVSGRLSMHKGIDFAAPAGSDVVA
VAAGVVTWSGRKNGYGNVVEVGHADGYTTLYAHNQKNLVEVGDLVKRGQVLAKVGSTG
RSTGYHVHFEVMKDGRVMNPQSFIARASVSE"

gene 5409. .5804
/locus_tag="PA4405"

CDS 5409. .5804
/locus_tag="PA4405"
/note="Protein name confidence: Class 4 (homologs of previously reported genes of unknown function, or no similarity to any previously reported sequences)"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07793.1"
/db_xref="GI:9950637"
/translation="MLFAEAQRICRLQQLLESQLQPAARPYCHVASWRGGCLLLIVTD
GHWATRLRYQQRRLQRQLQAFEEFANLAKILFKVQPTTSTQRAARRSLTLSAVAAENL
RESAEIVGNPRLREALERLASRAEKSRQE"

gene complement(5865. .6776)
/gene="lpxC"
/locus_tag="PA4406"
/note="synonyms: envA, asmB"

CDS complement(5865. .6776)
/gene="lpxC"
/locus_tag="PA4406"
/note="Protein name confidence: Class 1 (function experimentally demonstrated in P. aeruginosa); Alt protein name: EnvA protein"
/codon_start=1
/transl_table=11
/product="UDP-3-O-acyl-N-acetylglucosamine deacetylase"
/protein_id="AAG07794.1"
/db_xref="GI:9950638"
/translation="MIKQRTLKNIIRATGVGLHSGEKVYLTLKPAPVDTGIVFCRTDL
DPVVEIPARAENVGETTMSTTLVKGDVKVDTVEHLLSAMAGLGIDNAYVELSASEVPI
MDGSAGPFVFLIQSAGLQEQEAAKKFIRIKREVSVEEGDKRAVFVPFDGFKVSFEIDF
DHPVFRGRTQQASVDFSSTSFVKEVSRARTFGFMRDIEYLRSQNLALGGSVENAIVVD
ENRVLNEDGLRYEDEFVKHKILDAIGDLYLLGNSLIGEFRGFKSGHALNNQLLRTLIA
DKDAWEVVTFEDARTAPISYMRPAAAV"

gene complement(6889. .8073)
/gene="ftsZ"
/locus_tag="PA4407"

CDS complement(6889. .8073)
/gene="ftsZ"
/locus_tag="PA4407"
/note="Protein name confidence: Class 2 (high similarity to functionally studied protein)"
/codon_start=1
/transl_table=11
/product="cell division protein FtsZ"
/protein_id="AAG07795.1"
/db_xref="GI:9950639"
/translation="MFELVDNIAQTAVIKVIGVGGGGGNAVNHMAKNNVEGVEFICAN
TDAQALKNIAARTVLQLGPGVTKGLGAGANPEVGRQAALEDRERISEVLEGADMVFIT
TGMGGGTGTGAAPIIAEVAKEMGILTVAVVTRPFPFEGRKRMQIADEGIRALAESVDS
LITIPNEKLLTILGKDASLLAAFAKADDVLAGAVRGISDIIKRPGMINVDFADVKTVM
SEMGMAMMGTGCASGPNRAREATEAAIRNPLLEDVNLQGARGILVNITAGPDLSLGEY
SDVGNIIEQFASEHATVKVGTVIDADMRDELHVTVVATGLGARLEKPVKVVDNTVQGS
AAQAAAPAQREQQSVNYRDLDRPTVMRNQSHGSAATAAKLNPQDDLDYLDIPAFLRRQ
AD"

gene complement(8124. .9377)
/gene="ftsA"
/locus_tag="PA4408"

Query Match 99.6%; Score 1078.2; DB 1; Length 24000;
Best Local Similarity 99.7%; Pred. No. 4.7e-131;
Matches 1080; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy        1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17361 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 17302

Qy       61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17301 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 17242

Qy      121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
            |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    17241 GGGCCCTGGATGATCCGTACCTTGCAGATCCGCCAGATCGGCCAGGCCGTGCGCAACGAC 17182

Qy      181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 17122

Qy      241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17121 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 17062

Qy      301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
            |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    17061 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTGGACGACTACCGCAAGGTG 17002

Qy      361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    17001 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 16942

Qy      421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16941 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 16882

Qy      481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16881 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 16822

Qy      541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16821 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 16762

Qy      601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    16761 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 16702
```

```
Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16701 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 16642

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16641 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 16582

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16581 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 16522

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16521 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 16462

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16461 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 16402

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db      16401 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGAGCCGCGCGTGATCGTG 16342

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      16341 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 16282

Qy       1081 TGA 1083
              |||
Db      16281 TGA 16279
```

```
RESULT 3
AE016779
LOCUS       AE016779               301995 bp    DNA     linear   BCT 12-DEC-2002
DEFINITION  Pseudomonas putida KT2440 section 6 of 21 of the complete genome.
ACCESSION   AE016779 AE015451
VERSION     AE016779.1  GI:26557023
KEYWORDS    .
SOURCE      Pseudomonas putida KT2440
  ORGANISM  Pseudomonas putida KT2440
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 301995)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Complete genome sequence and comparative analysis of the
            metabolically versatile Pseudomonas putida KT2440
  JOURNAL   Environ. Microbiol. 4 (12), 799-808 (2002)
REFERENCE   2  (bases 1 to 301995)
  AUTHORS   Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,
            Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,
            Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,
            Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,
            Holtzapple,E., Scanlan,D., Tran,K., Moazzez,A., Utterback,T.,
            Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,
            Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,
            Duesterhoft,A., Tummler,B. and Fraser,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2002) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .301995
                     /organism="Pseudomonas putida KT2440"
                     /mol_type="genomic DNA"
                     /strain="KT2440"
                     /db_xref="taxon:160488"
     gene            89. .712
                     /gene="sspA"
                     /locus_tag="PP1320"
     CDS             89. .712
                     /gene="sspA"
                     /locus_tag="PP1320"
                     /note="similar to GP:9864180; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="stringent starvation protein A"
                     /protein_id="AAN66944.1"
                     /db_xref="GI:24982777"
                     /translation="MGATNRLACYSDPADHYSHRVRLVLAEKGVSVQLIDVDPAHLPR
                     KLAEVNPYGSVPTLVDRDLALYESTVVMEYLEERYPHPPLMPVYPVARGNSRLLMHRI
                     QRDWCALADTVLDPRSSEAARTEARKALRESLTGVSPLFSEFACFMSDEQSLVDCCLL
                     PILWRLPVLGIELPRQAKPLLDYMERQFAREPFQASLSSVEREMRKL"
     gene            693. .1154
                     /gene="sspB"
                     /locus_tag="PP1321"
     CDS             693. .1154
                     /gene="sspB"
                     /locus_tag="PP1321"
                     /note="similar to GB:S50157, GB:X60221, SP:P24539, and
                     PID:509291; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="stringent starvation protein B"
                     /protein_id="AAN66945.1"
                     /db_xref="GI:24982778"
                     /translation="MKCASFKEPLMNSSRPYLVRALYEWIVDNDCTPHMLVNAEFPKV
                     QVPDGFASDGQIVLNISPSAVRSLHMDNDAVSFEGRFSGVAHSLFVPVGAILGIYARE
                     NGQGMVFELEPSSMDGDELEDDDVQPDDDGPPDDGGQPPRPTGRPSLKVVK"
     gene            complement(1216. .1794)
                     /locus_tag="PP1322"
     CDS             complement(1216. .1794)
                     /locus_tag="PP1322"
                     /codon_start=1
                     /transl_table=11
                     /product="lipoprotein, putative"
                     /protein_id="AAN66946.1"
                     /db_xref="GI:24982779"
                     /translation="MTPMRLGLMALTLCLSVTGCSSVLTSTRNSPIEDDRGTRTIGSK
                     IDDSLIETKAAVNISKANPNLDKGSHIVVSSYNGIVLLAGQTPRADLKSLAEQTASQV
                     QRVKKVHNELQVMQPSSILARNNDAWLTTKIKTQMLTDSAVPSSRIKVITENGIVYLL
                     GLVTRQEANSATAVVQGVSGVQKIVKLFEYID"
     gene            complement(1791. .2378)
                     /gene="gmhA"
                     /locus_tag="PP1323"
     CDS             complement(1791. .2378)
                     /gene="gmhA"
                     /locus_tag="PP1323"
                     /note="identified by match to PFAM protein family HMM
                     PF01380"
                     /codon_start=1
                     /transl_table=11
                     /product="phosphoheptose isomerase"
                     /protein_id="AAN66947.1"
                     /db_xref="GI:24982780"
                     /translation="MQSRIRRLFEASIETKQQAMDILAPHIEQASLVMVNALLNEGKM
                     LACGNGGSAGDAQHFSSELLNRFERERPSLPAIALTTDSSTLTSIANDYSYNEVFSKQ
                     IRALGQPGDVLLAISTSGNSANVIQAIQAAHDREMIVVALTGRDGGGMASLLLPEDVE
                     IRVPSTVTARIQEVHLLAIHCLCDLIDSQLFGSEE"
     gene            complement(2471. .2845)
                     /locus_tag="PP1324"
     CDS             complement(2471. .2845)
                     /locus_tag="PP1324"
                     /note="identified by match to TIGR protein family HMM
                     TIGR00252"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein TIGR00252"
```

/protein_id="AAN66948.1"
/db_xref="GI:24982782"
/translation="MMAAASPTRAGQAAETQALEYLQGQGLQLLARNWRCKGGELDLV
MLDADTVVFVEVRYRLHAGFGGALDSIDGRKQKRLVLAATLFLQKEAHWGNYPCRFDV
VALQGSHHAGRPLQWLKNAFEC"
gene complement(2842. .4659)
/locus_tag="PP1325"
CDS complement(2842. .4659)
/locus_tag="PP1325"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
/protein_id="AAN66949.1"
/db_xref="GI:24982783"
/translation="MIACLRLLTALCLAALLAACASSPSSSLGELPRTPDASIEQLLE
KAASSKSAEDAALLRLSAADLAYKQKDFPRAARILEQVPLDTLKPAQQVFASTLAAEL
AMSRNQPKAALTALAHPSLQRLAELPADLQARTYSVHAAALEADGQALAAAQQRILLA
PLLGGQAASTNNDAIWALVASLPAEQLQQPTNDQTLAGWTSLAFAVKSAGTLEQQQAA
IEAWVKQHPDHPAAQQLPLALTKLKELASQPLTKIALLLPQEGPLAGVARALRDGFMA
AHFQAQQSGQPAPAVQVFDSSRIGSLDDFYRQAQAAGVQLVIGPLEKPLVKQLAAKPQ
LPITTLALNYADAGQKAPPQLFQFGLAAEDEAREVARRARADGMVRAVALVPSGEWGD
RVLAAFRQDWEGNGGTLLAAERIAQPVALAQQIADLFQLRQSEGRAKSLQSTVGGSIA
AQPSRRQDIDFIFLASTPQQAQQIKPTLNFQYAGDVPVYATSNLYSASGDINQYNDMN
GIRFCETPWLLDTSNSLRQQVVQQWPQAAGSLGRLYAMGVDAYSLAPRLGQLKALPDN
RVLGLSGSLSINANQRVERQLPWAEFSGGQVKRLPDTVR"
gene 4837. .5721
/locus_tag="PP1326"
CDS 4837. .5721
/locus_tag="PP1326"
/note="identified by match to PFAM protein family HMM
PF00590"
/codon_start=1
/transl_table=11
/product="tetrapyrrole methylase family protein"
/protein_id="AAN66950.1"
/db_xref="GI:24982784"
/translation="MCAVTDVAGASKSTVGTLYVVATPIGNLDDMSARALKVLADVAL
IAAEDTRHSIRLLQHFGIDTPLAACHEHNERDEGGRFLSKLLAGENVALVSDAGTPLI
SDPGYHLVRQARAAGVSVVPVPGACALIAALSAAGLPSDRFIFEGFLPAKQAGRRARL
EQVREEPRTLIFYEAPHRILECLEDMEAVFGGDRPALLARELTKTFETLKGLPLAELR
AFVAGDSNQQRGECVVLVGGWSAPEGEQAISAEAQRVLDLLLAELPLKRAAALAAEIT
GVRKNLLYQLALEKQKAE"
gene complement(5744. .6164)
/locus_tag="PP1327"
/note="This region contains a gene with one or more
premature stops or frameshifts, and is not the result of a
sequencing artifact; conserved hypothetical protein,
degenerate; similar to GB:X16944, and SP:P15488;
identified by sequence similarity; putative"
gene 6455. .6928
/locus_tag="PP1328"
CDS 6455. .6928
/locus_tag="PP1328"
/note="identified by match to PFAM protein family HMM
PF02381"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00242"
/protein_id="AAN66951.1"
/db_xref="GI:24982786"
/translation="MWGITAVFRGANAVSLDAKGRLAMPSRYRDELDSRCNGQLIVTI
DAVDPCLCVYPLDEWEQIEAKLRALPSLREENRRLQRLLIGNAVDLELDGSGRFLVPP
RLREYAKLDKKAMLVGQLNKFQLWDEDAWNAVSAADLAAIQQPGAMPDELRDLIL"
gene 6925. .7872
/locus_tag="PP1329"
CDS 6925. .7872
/locus_tag="PP1329"
/note="similar to GP:9950652; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein TIGR00006"
/protein_id="AAN66952.1"
/db_xref="GI:24982787"
/translation="MTIDSGFNHITVLLDEAVEALALRADGCYLDGTFGRGGHSRLIL
SKLGPQGRLLGFDKDPQAIATGQALAAEDGRFVIVQRSFAELGAEVAARGLHGKVSGV
LLDLGVSSPQLDDPERGFSFLNDGPLDMRMNPDQGVSAAEFIATAPVEEIARVFKEYG
EERFAGRMARAVVERREKQPFTRTADLAEVLKVANPAWEKGKNPATRAFQGLRIHVNN
ELGDLEAGLEAALDALEVGGRLAVISFHSLEDRIVKLFMRKLVKGEADNLPRNLPVQH
KVFEPKIKLIGKAQFASEAELKANPRSRSAVMRVAEKLR"
gene 7869. .8162
/gene="ftsL"
/locus_tag="PP1330"
CDS 7869. .8162
/gene="ftsL"
/locus_tag="PP1330"
/note="similar to GB:M11186, GB:M62611, SP:P01178,
PID:189411, PID:386991, and PID:825699; identified by
sequence similarity; putative"

Query Match 75.7%; Score 820.2; DB 1; Length 301995;
Best Local Similarity 84.9%; Pred. No. 8.6e-98;
Matches 918; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

```
Qy        1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
            |||||  ||||||  |||||||  ||  ||||||||||||  ||||  |||||||  |||||  |||
Db    12754 ATGCTGCTGCTGTTGGCCGAGTATCTGCAACAGTTCCACAAAGGCTTCGCGGTCTTTCAG 12813

Qy       61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
            ||||||  ||||||||||||  |||||  |  ||  ||  ||||||  ||||  |||  ||||||||||
Db    12814 TACCTGTCCCTGCGCGGGATTCTTGGTGTACTGACCGCGTTGTCCCTGGCGCTGTGGCTG 12873

Qy      121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
            ||  ||||||||||||||||||  ||||  ||  |  ||||||||||||||||||||||  ||||||
Db    12874 GGCCCCTGGATGATCCGTACCCTGCAAATTCGCCAGATCGGCCAGGCCGTGCGTAACGAC 12933

Qy      181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
            ||  |||||  ||||||||||  |||  ||||||||||||||||||  ||  ||  ||||||||
Db    12934 GGCCCGCAATCGCACCTGTCCAAGTCCGGCACCCCGACCATGGGTGGGGCGCTGATCCTG 12993

Qy      241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              |  |||||  |||  |||||||  ||||||||||  ||  ||  |||||||||  ||  |||||
Db    12994 TCGGCCATCGCCGTCAGCACCTTGCTGTGGGCCGACCTGAGCAACCGCTATGTATGGGTT 13053

Qy      301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
            ||||||  |  ||  ||||||  ||||||||||||  ||||||||  ||  ||||||||||||||||
Db    13054 GTGCTGATTGTCACCCTGGCGTTCGGTGCCATTGGCTGGGTCGATGACTACCGCAAGGTG 13113

Qy      361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
            |||||  ||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db    13114 ATCGAAAAGAACTCGCGTGGCCTTCCGAGCCGCTGGAAGTACTTCTGGCAATCGGTGTTC 13173

Qy      421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
            |||  |  ||  ||  ||||||||  ||||  |||  ||  ||||  ||||||||||  |||
Db    13174 GGCCTGGCGGCGGCGGTGTTCCTGTACAAGACGGCGCCAACCAGCGTCGAGACCACGCTG 13233

Qy      481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
            |||  |||||  |  |  |||  |||  ||  |  |||||||  ||  ||||||  ||  ||||||
Db    13234 ATCCTGCCGTTCATCAAGGATGTCACCATTCCGTTGGGCGTCGGCTTCGTCGTACTGACC 13293

Qy      541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
            ||||||||||  ||||||||||||||  |||||  ||  ||||||||  ||  ||  |||||||||
Db    13294 TACTTCGTCATTGTCGGCTCGAGCAACGCGGTCAATCTCACCGATGGCCTTGACGGCCTG 13353

Qy      601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
            ||  ||||||||||||||  |||||  |  ||||||||||||||||||||||||||||||||||||
Db    13354 GCAATCATGCCGACGGTGATGGTCGGCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGT 13413

Qy      661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
            |||||||||||||||||  ||||||||||  |||  ||||  ||||||  |  |||||  ||||||
Db    13414 AACGTGAAGTTCGCCGAATACCTGCTGATCCCCTACGTGCCGGGCTCGGGCGAACTGATC 13473

Qy      721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
            |||||||||  |||||||  ||||  ||||||||  |||||  ||  ||||||||||||  ||||||
Db    13474 GTGTTCTGCGGCGCGCTGATCGGTGCCGGCCTGGGCTTTCTGTGGTTCAACACTTATCCG 13533
```

```
Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||| || ||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db      13534 GCGCAAGTGTTCATGGGCGACGTCGGCGCGCTGGCACTGGGCGCCGCGCTGGGCACCATC 13593

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || || |||||||||||||| |||||||||||||||||||| ||   | |||||  |||||
Db      13594 GCCGTCATCGTGCGCCAGGAAATCGTGCTGTTCATCATGGGCGGCATCTTCGTGGTGGAA 13653

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||| |||||| |||||||||||| ||||||||||||||||    ||||| |||||
Db      13654 ACCCTGTCGGTGGTGATCCAGGTCGCCTCCTTCAAGCTGACCGGCAAGCGCGTGTTCCGC 13713

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||| |||||||| || ||||| || |||||||| || ||  | ||||||||
Db      13714 ATGGCGCCGATTCATCACCACTTTGAACTCAAGGGCTGGCCAGAGCCCAGGGTGATCGTC 13773

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              || |||||||||||||||||||||||| ||||| ||||| || ||| |||| ||| |
Db      13774 CGTTTCTGGATCATCACCGTGATCCTGGTACTGATTGGCCTTGCAACCCTGAAACTGAGG 13833

Qy       1081 T 1081
              |
Db      13834 T 13834
```

RESULT 4
AE016871/c

```
LOCUS       AE016871              311600 bp    DNA     linear   BCT 09-OCT-2003
DEFINITION  Pseudomonas syringae pv. tomato str. DC3000 section 16 of 21 of the
            complete genome.
ACCESSION   AE016871 AE016853
VERSION     AE016871.1  GI:28854552
KEYWORDS    .
SOURCE      Pseudomonas syringae pv. tomato str. DC3000
  ORGANISM  Pseudomonas syringae pv. tomato str. DC3000
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE   1  (bases 1 to 311600)
  AUTHORS   Buell,C.R., Joardar,V., Lindeberg,M., Selengut,J., Paulsen,I.T.,
            Gwinn,M.L., Dodson,R.J., Deboy,R.T., Durkin,A.S., Kolonay,J.F.,
            Madupu,R., Daugherty,S., Brinkac,L., Beanan,M.J., Haft,D.H.,
            Nelson,W.C., Davidsen,T., Zafar,N., Zhou,L., Liu,J., Yuan,Q.,
            Khouri,H., Fedorova,N., Tran,B., Russell,D., Berry,K.,
            Utterback,T., Van Aken,S.E., Feldblyum,T.V., D'Ascenzo,M.,
            Deng,W.L., Ramos,A.R., Alfano,J.R., Cartinhour,S., Chatterjee,A.K.,
            Delaney,T.P., Lazarowitz,S.G., Martin,G.B., Schneider,D.J.,
            Tang,X., Bender,C.L., White,O., Fraser,C.M. and Collmer,A.
  TITLE     The complete genome sequence of the Arabidopsis and tomato pathogen
            Pseudomonas syringae pv. tomato DC3000
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (18), 10181-10186 (2003)
   PUBMED   12928499
REFERENCE   2  (bases 1 to 311600)
  AUTHORS   Buell,R., Joardar,V., Khouri,H., Fedorova,N., Tran,B., Russell,D.,
            Berry,K., Utterback,T., Van Aken,S., Feldblyum,T., Gwinn,M.,
            Dodson,R., DeBoy,R., Durkin,A., Kolonay,J., Madupu,R.,
            Daugherty,S., Brinkac,L., Beanan,M., Haft,D., Selengut,J.,
            Nelson,W., Davidsen,T., White,O., Fraser,C. and Collmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2003) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1. .311600
                     /organism="Pseudomonas syringae pv. tomato str. DC3000"
                     /mol_type="genomic DNA"
                     /strain="DC3000"
                     /db_xref="taxon:223283"
                     /note="pathovar: tomato"
     gene            131. .601
                     /gene="bfr"
                     /locus_tag="PSPTO4160"
     CDS             131. .601
                     /gene="bfr"
                     /locus_tag="PSPTO4160"
                     /note="similar to SP:P22759, and SP:P22759; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="bacterioferritin"
                     /protein_id="AAO57616.1"
                     /db_xref="GI:28854553"
                     /translation="MKGDITVIQHLNKILGNELVAINQYFLHARMYEDWGLKKLGGHE
                     YKESIHAMKYADKLIKRTLFLEGLPNLQDLGKILIGEHTKEMLECDLKIEQKGLADLK
                     AAIAHFETVGDFGSRELLEDILEPKEEQIDWLETQLGLIVKVGIENYLQSQMGD"
     gene            complement(689. .1015)
                     /locus_tag="PSPTO4161"
     CDS             complement(689. .1015)
                     /locus_tag="PSPTO4161"
                     /note="identified by match to TIGR protein family HMM
                     TIGR00365"
                     /codon_start=1
                     /transl_table=11
                     /product="glutaredoxin-related protein"
                     /protein_id="AAO57617.1"
                     /db_xref="GI:28854554"
                     /translation="MDIIETIKEQIASNTILLYMKGAPNAPQCGFSAKASQALMACGE
                     KFAYVDILQNPEIRANLPKYANWPTFPQLWVAGELVGGSDIITEMMADGSLQTLVKEA
                     SAAKAE"
     gene            complement(1104. .3302)
                     /locus_tag="PSPTO4162"
     CDS             complement(1104. .3302)
                     /locus_tag="PSPTO4162"
                     /note="similar to GP:5441785; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="oxidoreductase, molybdopterin-binding"
                     /protein_id="AAO57618.1"
                     /db_xref="GI:28854555"
                     /translation="MARVIARHHSHCSAIHRIAMTRPEAAMTRTQHLRACHLCEAICG
                     LAIETVTEPGAAPKITSIKGDPLDTFSRGHICPKAVALQDIQNDPDRLRQPMLRTGDQ
                     WQPIEWQAAFDLVAERLYAIQQQHGQNAVAVYQGNPSVHNYGLMTHSNYFLGLLKTRN
                     RFSATSVDQLPHHLTSFLMYGHGMLLPIPDIDHTDFMLILGGNPLASNGSIMTVPDVE
                     KRLKAIQQRGGKLVVVDPRRSETAAIADQHVFVRPGGDAALLFGLLNTLFEEGLTRES
                     HLPVDGLDHVRHSIAGFSAEAMSPRCGIAAVQIRQLARDFAGADTALCYGRMGVSTQA
                     FGTLCHWLAQLINLVTGNLDRVGGTLCTEPAVDLVSSTSGGHFNVWQSRVSGLPEYGG
                     ELPVSALAEEMLVEGQGQVRALVTVAGNPVLSTPNGRQLDQALEGLEFMLSIDLYINE
                     TTRHADLILPSTSALENDHYDTTFNTLAVRNVTRFNRAIFDKPEGALHDWEIFVGLAA
                     SFAAKAQRELKPTVPPAQMIDRVLRAGRYGAASPHNLSLETLASHPHGMDLGALKPNL
                     TDRLKTANGRIQAAPEVIMADLVRFAADAGPRFGEKAGQLLLIGRRHVRSNNSWMHNY
                     HRLVKGKPRHQLLMHPDDLSHRGLSDGQQVRVSSRVGVIEVEVLGSLDMMPGVVSLPH
                     GWGHSRSGVKMEIARNQPGVSANDLTDERQLDELSGNAALNGVPVQVASC"
     gene            complement(3357. .3650)
                     /locus_tag="PSPTO4163"
     CDS             complement(3357. .3650)
                     /locus_tag="PSPTO4163"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAO57619.1"
                     /db_xref="GI:28854556"
                     /translation="MDSKEHQLTELLGLTARTLTHLTASMTSMSFELLRSEDEVTRNA
                     GRRMIDRMATISSGLDEHWRLIGDLTGIHMAQEQVETVTEIQLQRKIVTPIGG"
     gene            4100. .5020
                     /gene="argF"
                     /locus_tag="PSPTO4164"
     CDS             4100. .5020
                     /gene="argF"
                     /locus_tag="PSPTO4164"
                     /note="identified by match to PFAM protein family HMM
                     PF02729"
                     /codon_start=1
                     /transl_table=11
                     /product="ornithine carbamoyltransferase"
                     /protein_id="AAO57620.1"
                     /db_xref="GI:28854557"
```

/translation="MNARHFLSMMDYTPDELLGLIRRGVELKDLRNRGVLFEPLKNRV
LGMIFEKSSTRTRLSFEAGMIQLGGQAIFLSHRDTQLGRGEPIADSAKVISRMLDAVM
IRTYAHDNLTEFAANSRVPVINGLSDDLHPCQLLADMQTFLEHRGSIKGKTVAWIGDG
NNMCNSYIEAAIQFDFQLRVACPAGYEPNPEFLALAGERVTVVRDPKAAVAGAHLVST
DVWTSMGQEEETARRKALFAPFQVTRALLDLADKDVLFMHCLPAHRGEEISVDLLDDA
RSVAWDQAENRLHAQKALLEFLVAPCFQPA"

gene 5026. .6087
/locus_tag="PSPTO4165"

CDS 5026. .6087
/locus_tag="PSPTO4165"
/note="similar to GB:Z48042, SP:Q14444, and PID:662994;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ABC transporter, ATP-binding protein"
/protein_id="AAO57621.1"
/db_xref="GI:28854558"
/translation="MLLNLRNLACGYQNQRVVQDLNLHLNAGDIGCLLGSSGCGKTTT
LRAIAGFEPIHAGEITLAGDVISRPGWTLAPEKRRIGMVFQDYALFPHLSVADNIAFG
IRQHPRLDQVIGELLELVNLGALGKRYPHELSGGQQQRVALARALAPEPQLLLLDEPF
SNLDGELRRRLSHEVRDILKARGTSAILVTHDQEEAFAVSDHVGVFKEGRLEQWDTPY
NLYHEPQTPFVASFIGQGYFIRGQMIEPESVHTELGVLRGNRAYPGVRGSAVDVLLRP
DDIVYAPDSDLKARVIGRTFKGASTLYRLQLPTGSQLEAIFPSHADHRQGEEVGIRVA
ADHLVLFAAQGSVAAYAAG"

gene complement(6096. .6566)
/locus_tag="PSPTO4166"

CDS complement(6096. .6566)
/locus_tag="PSPTO4166"
/note="similar to SP:P21484; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ybaK/ebsC protein"
/protein_id="AAO57622.1"
/db_xref="GI:28854559"
/translation="MTPALDLLKKNRAEHRIHSYEHDPKAASYGLEAAEKLGLDAARV
FKTLLASTEKGELLVAVVPVVGTLDLKALANAAGAKKTEMADPAAAQRSTGYLLGGIS
PLGQKKRLRTFIDETAQLFESIYVSAGRRGLEVELTAAVLAEHTQARFAPIGRE"

gene 6918. .7775
/gene="glpF"
/locus_tag="PSPTO4167"

CDS 6918. .7775
/gene="glpF"
/locus_tag="PSPTO4167"
/note="similar to SP:P11244; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glycerol uptake facilitator protein"
/protein_id="AAO57623.1"
/db_xref="GI:28854560"
/translation="MTIALKQPTLTGQCVAEFLGTALMIFFGTGCVAALKVAGATFGL
WEICIIWGMAVSMGIYLSAGISGAHLNPAVSIALSLFAGFEKRKLPFYISAQIAGAFC
GAGLVYLLYISLFFDFEHAHHIIRGSEQSLELASVFSTYPNPAISVGQAFLVEVVITT
ILMGVIMALGDDNNGLPRGALAPLLIGLLVAVIGSAMGPLTGFAMNPARDFGPKLMTF
FAGWGEIAFTGGRDIPYFLVPIFAPILGACLGAAGYRALIARHLPSTVTVENEKEAPV
VRGKVQASS"

gene 7805. .9310
/gene="glpK"
/locus_tag="PSPTO4168"

CDS 7805. .9310
/gene="glpK"
/locus_tag="PSPTO4168"
/note="identified by match to PFAM protein family HMM
PF02782"
/codon_start=1
/transl_table=11
/product="glycerol kinase"
/protein_id="AAO57624.1"
/db_xref="GI:28854561"
/translation="MTDTQNKNYIIALDQGTTSSRAIIFDRDANVVSTAQSEFVQHYP
QAGWVEHDPMEIFATQTACMTKALAQADLHHDQIAAIGITNQRETTVIWERDTGRPIY
NAIVWQCRRSTEICQQLKRDGLEEYIKDTTGLVLDPYFSGSKVKWILDHVEGSRERAR
KGELMFGTIDTWLIWKFTGGKVHVTDYTNASRTMLFNIHTLEWDQRMLDVLDIPREIL
PEVKSSSEVYGHSKSGIPIAGIAGDQQAALFGQMCVEPGQAKNTYGTGCFLLMNTGKK
AVKSAHGMLTTIGCGPRGEVAYALEGAVFNGGSTVQWLRDELKLINDALDTEYPAGKV
KDSNGVYLVPAFTGLGAPYWDPYARGALFGLTRGVKVDHIIRAALESIAYQTRDVLDA
MQQDSGERLKSLRVDGGAVANNFLMQFQADILGTHVERPQMRETTALGAAFLAGLAIG
FWSSLDELRNKAVIERVFEPACEEAHREKLYAGWQKAVARTRDWEPHENEE"

gene 9421. .10176
/gene="glpR"

Query Match 70.9%; Score 767.4; DB 1; Length 311600;
Best Local Similarity 81.9%; Pred. No. 6e-91;
Matches 885; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

```
Qy        1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
Db   292940 ATGCTGCTGCTGTTGGCCGAGTTCTTACAACAGTTCTACAAAGGCTTTGCGGTCTTTCAG 292881

Qy       61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
Db   292880 TACCTGTCTCTGCGCGGGATCCTTGGCGTGCTCACTGCGCTGACGTTGTCGCTGTGCCTG 292821

Qy      121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
Db   292820 GGGCCGTGGATGATCCGCACCCTGCAGATGCGCCAGATCGGCCAGTCGGTCCGTAACGAT 292761

Qy      181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
Db   292760 GGTCCGCAATCGCACTTGTCCAAGTCCGGTACGCCGACCATGGGCGGTGCGCTGATCCTC 292701

Qy      241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
Db   292700 TCGTCCATCGGTATCAGCACCTTGCTCTGGGCTGATTTGAGCAACCGTTACGTCTGGGTT 292641

Qy      301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
Db   292640 GTGTTGCTGGTGACCTTTCTGTTCGGTGCCATCGGCTGGGTTGATGACTACCGCAAGGTC 292581

Qy      361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db   292580 ATTGAAAAGAATTCCCGTGGTCTGCCGAGTCGCTGGAAGTATTTCTGGCAGTCGGTGTTC 292521

Qy      421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db   292520 GGCCTTTGCGCGGCAATCTTTCTGTACACGACTGCGCCATCGGCGACCGAAACCACATTG 292461

Qy      481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db   292460 ATCGTGCCGATGCTCAAGGATGTGCGCATTCCGCTCGGCATCGGCTTCATCGTGCTGACG 292401

Qy      541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db   292400 TACTTTGTGATTGTCGGTTCCAGCAACGCAGTCAACCTGACTGACGGCCTGGACGGGCTG 292341

Qy      601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db   292340 GCGATCATGCCGACGGTGATGGTCGGCGGTGCGCTGGGCATCTTCTGCTACCTGTCGGGT 292281

Qy      661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db   292280 AACGTGAAATTCGCGGAATACCTGCTGATTCCCTACGTGCCGGGGGCGGGTGAGCTGATT 292221

Qy      721 GTGTTCTGCGCCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db   292220 GTCTTCTCGGGTGCACTGATTGGTGCGGGGCTCGGGTTTCTCTGGTTCAACACCTATCCG 292161

Qy      781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db   292160 GCTCAGGTGTTCATGGGCGACGTCGGTGCGCTGGCACTGGGCGCTGCGCTGGGCACCATG 292101

Qy      841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db   292100 GCCGTGATCGTGCGTCAGGAAATGGTCCTGTTCATCATGGGCGGCGTGTTCGTGATGGAA 292041
```

```
Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||||| || ||| | || ||||| || || ||||| |||||||| |||||||| |||||
Db     292040 ACCCTTTCAGTGGTCATTCAGGTTGCGTCTTTCAAACTGACCGGCCGCCGCGTATTCCGC 291981

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||||| || ||||| || ||||| |||||||||| || || ||||||||||
Db     291980 ATGGCGCCGATTCACCACCACTTTGAACTCAAAGGCTGGCCCGAGCCACGCGTGATCGTC 291921

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              || |||||||||||||||||| |||||||| |||||||| |||||  |||| ||| |
Db     291920 CGTTTCTGGATCATCACCGTGATTCTGGTGCTCATCGGCCTTGCCACACTGAAACTGAGG 291861

Qy       1081 T 1081
              |
Db     291860 T 291860
```

RESULT 5
AE012171

LOCUS	AE012171 13045 bp DNA linear BCT 23-MAY-2002
DEFINITION	Xanthomonas campestris pv. campestris str. ATCC 33913, section 79 of 460 of the complete genome.
ACCESSION	AE012171 AE008922
VERSION	AE012171.1 GI:21111728
KEYWORDS	.
SOURCE	Xanthomonas campestris pv. campestris str. ATCC 33913
ORGANISM	Xanthomonas campestris pv. campestris str. ATCC 33913 Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales; Xanthomonadaceae; Xanthomonas.
REFERENCE	1 (bases 1 to 13045)
AUTHORS	da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
TITLE	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
JOURNAL	Nature 417 (6887), 459-463 (2002)
MEDLINE	22022145
PUBMED	12024217
REFERENCE	2 (bases 1 to 13045)
AUTHORS	da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil

FEATURES Location/Qualifiers

source	1. .13045 /organism="Xanthomonas campestris pv. campestris str. ATCC 33913" /mol_type="genomic DNA" /strain="ATCC 33913" /db_xref="ATCC:33913" /db_xref="taxon:190485" /note="pathovar: campestris"
gene	189. .644 /gene="XCC0717"
CDS	189. .644 /gene="XCC0717" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAM40032.1" /db_xref="GI:21111729" /translation="MHSVFQGETAITVDDKGRMAVPTAYRDLVARVSGNRLVLTYNPF EAGCLWLYAEKEWERVRDDVMSKPNTQRVVRTLQQKLVGSSAVLELDANGRLSIPASH RNAVGIEKKAVLLGMGDKFELWSEQAHRALIQQTLSDGDLGDELLDLRL"
gene	689. .1717 /gene="mraW" /note="synonym: XCC0718"
CDS	689. .1717 /gene="mraW" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="cell division protein" /protein_id="AAM40033.1" /db_xref="GI:21111730" /translation="MSQPPAAHVPVLYTQVLEGLQVTENGTYLDGTFGRGGHARGVLE HLGPGGRLLVMDKDPEAIAVAEHTFGGDARVSIHRGSFAGLGQVVAAATVDGILLDLG VSSPQLDVAGRGFSFGKDGPLDMRMDPDSGQSAAGWLAQATDREIADVLWTYGEERQS RRIARAIVARRGEQPLLRTAQLADLIASVMPRGDSKTHPATRSFQAIRIHINRELADL EAGLDAALGALKPGGRLAVISFHSLEDRIVKQFMARYAKAPPSNRRLPEAQPFVPTLQ LVSGAIKADDSELAVNPRARSAVLRVAEKLGMENRESGMGKGHGAAASRFPTPDSRFP TSPNGDAP"
gene	1714. .1977 /gene="ftsL" /note="synonym: XCC0719"
CDS	1714. .1977 /gene="ftsL" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="cell division protein" /protein_id="AAM40034.1" /db_xref="GI:21111731" /translation="MSRLLLIVLLACSIASAIGVVYMRHMHRKLFVQLSKLEHSRDEL NIEFGRLQLEQATWAESNRVDQVSRERIGMKFPETSDIVVIRP"
gene	2024. .3895 /gene="ftsI" /note="synonym: XCC0720"
CDS	2024. .3895 /gene="ftsI" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="penicillin-binding protein 3" /protein_id="AAM40035.1" /db_xref="GI:21111732" /translation="MCAMKAGRNRPRSNFNLRGRLTLVGVALGLCSVTLIGRAAYVQL VNRDFYQRQGEARYLRELPIATSRGMITDRNGEPLAVSTPVESIWVNPQELLRNPDRI PELATALGQPLDELNAKLAQKAGKEFMYLKRRINPDKAHAIVDLKIPGVFSQREFRRF YPQGEAMAHVLGFTNIDDRGQEGLELAFDEWLRGKPGSKKVVRDARGAIVESVDLVRP AQPGKDLTLSIDRRIQFLAYKELRNALVENKAAGGSMVVMDVATGEVLAMVNLPTYNP NSVTGINPSARRNRAVTDLVEPGSTMKPLTVATALTAGVVTKDTIIDTNPGYMAVGRF

```
                TIRDVPRNNGVLNVTGVITRSSNIGAAKIAAKVPDQTFYQSIRNFGYGSAPQSGFPGE
                SSGMVLQPQKWSGPSKTTMSYGYGLSVTPLQIAQAYATLGNGGKLTPPTFVRGQHNET
                RQVVTPEVARQVIDMMETVVTQGGAKGAAILGYHVAGKTGTARKNGAGGYERGHYNAL
                FAGLVPATRPRFATVIVINDPQGKVYYGGLVSAPVFHRVMEGALRLMDVPPDDIQSWL
                AAQAAGKTGQPVVKPQPADLDPALVPDPVDEVSAGIPTALASGATPAQPAPLQESRQ"
gene            3892. .5379
                /gene="murE"
                /note="synonym: XCC0721"
CDS             3892. .5379
                /gene="murE"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="UDP-N-acetylmuramoylalanyl-D-
                glutamate-2,6-diaminopimelate ligase"
                /protein_id="AAM40036.1"
                /db_xref="GI:21111733"
                /translation="MSRAMALSQLLPDVALARDVQVSGLVMDSRAVRPGDAFVAIAGF
                GAHGLGFAAQALASGASAILFEPPAPADLPVPADAIAVPGLRARLGVLADHFHGAPSQ
                AMRMIGVTGTNGKTSTVQLLAQALTLLGTPTGTIGTLGVGLYGAAVPTGFTTPLVLQT
                HAELAQLRDAGAQAVAMEVSSHALDQGRVDAVQFDVAVFTNLTRDHLDYHGDMAQYGA
                AKAKLFARAGLRAAVVNLDDAFGRTLFAALDPALHAIGVSSRAQAGATVQAQDLQLDH
                HGIQFTLHIGEAAHPVRSPLLGRFNVDNLLAVAGALHALDIAPAQIAEVLGRLQPIHG
                RMNRLGGAHGAPLVVVDYAHTPDALEQALTSLRSHAQDRLICVFGCGGERDTGKRPQM
                AAIAEVTADVAIVTDDNPRGEDGDVIVADILRGFARPDAAIVQRDRAAAIHQAISMAG
                ADDIVLIAGKGHEPYQEVAGVRHAFDDAAVAAQALLPRTGLGVRV"
gene            5376. .6860
                /gene="murF"
                /note="synonym: XCC0722"
CDS             5376. .6860
                /gene="murF"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="UDP-N-acetylmuramoylalanyl-D-glutamyl-2,
                6-diaminopimelate-D-alanyl-D-alanyl ligase"
                /protein_id="AAM40037.1"
                /db_xref="GI:21111734"
                /translation="MKLTPLSLIAHFAGGELHGDDAMIDAVGNDTRALASGSLYVALR
                GERFDGHDFAADAVARGASGLLVDRLLPALHVPQVVVANTELALARIAAGMQRDRATR
                VIALTGSNGKTSVKALLLAILAEAGRVDGITAYANPGNRNNEIGLPLAVIAAPDDADL
                AVYEMGAGKPGDIAYLTDIARPDVALVNNIAPAHLERMGSLLGVAQTKGAIYTALPAD
                GVAVINADDAFGAWFEQQLLAQPLQGRRVLRYGLHASAEITARQLRLHAAGAQFVLAT
                PQGEAPVVLPLPGRHNVLNALAATGLALGAGIALPIIAAGLAQATQVAGRQIAHTLAT
                GAVLIDDSYNANPGSLDAAIDALAAAPGEGWLVLGDMRELGPEAPALHAAAGRRARNA
                KLARLYALGALSAAAAQAFGDGGRVFATHAALAAALQADLAAAASQTQTQDSSLMHEH
                TGAVARAAASVAPPTLLVKGSRGSAMDRIVAALLAPSEDASHAA"
gene            6850. .7935
                /gene="mraY"
                /note="synonym: XCC0723"
CDS             6850. .7935
                /gene="mraY"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="phospho-N-acetylmuramoyl-pentapeptide-
                transferase"
                /protein_id="AAM40038.1"
                /db_xref="GI:21111735"
                /translation="MLLELARWLQQLESLFGLFNYLTFRGILAALTALFLSLWMGPAV
                IRKLAQFKGGQPIRQDGPQTHFSKAGTPTMGGSLILLTVTLSVLLWGDLRNRYVWLVL
                AVMICFGAIGWYDDWIKIVRRDPNGLKSRWKYLLQSIFGLAAGLFLYYTADVPAAITF
                YIPMFKAIALPLAGVSFVVIAYFWIVGFSNAVNLTDGLDGLAIMPTVLVACALGVFAY
                ASGNVVFAEYLKIPLIPGAGELIIICSAIAGAGLGFLWFNTYPAMVFMGDIGALSLGA
                VLGTVAVIVRQELVLVIMGGVFVIETLSVMIQVASFKLTGKRVFRMAPIHHHFELKGW
                PEPRVIVRFWIISVVLVLIGLATLKVR"
gene            7935. .9299
                /gene="ftsW"
                /note="synonym: XCC0724"
CDS             7935. .9299
```

```
Query Match             55.2%;  Score 598;  DB 1;  Length 13045;
Best Local Similarity   72.6%;  Pred. No. 9.9e-69;
Matches  788;  Conservative    0;  Mismatches  295;  Indels    3;  Gaps    1;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||| ||   |||||||   |  || || ||  |  | |     |||||  | ||| |
Db       6850 ATGCTGCTTGAGCTGGCCCGTTGGCTACAGCAACTGGAGAGCCTGTTCGGGCTGTTCAAC 6909

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||| | ||||||||  ||     | ||| ||||||||||  ||||||  ||||| ||
Db       6910 TATCTGACCTTCCGCGGCATCCTGGCAGCGCTGACCGCGCTGTTCCTGTCGTTGTGGATG 6969

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||    | ||||  || |   ||  |      | ||  |||||||  |  | ||| | |||
Db       6970 GGCCCGGCGGTGATTCGCAAGCTGGCGCAGTTCAAGGGCGGCCAGCCGATTCGCCAGGAC 7029

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || |||||| ||||  | ||  |||   ||||| |||||||||||  ||| | ||||||||
Db       7030 GGCCCGCAGACGCATTTTTCCAAGGCCGGCACGCCGACCATGGGTGGCTCGCTGATCCTG 7089

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              | | |  || |     | |||||||||  |||||   |||||||| ||||||  |
Db       7090 CTCACGGTCACCTTGTCGGTGTTGCTGTGGGGCGATCTGCGCAACCGCTATGTGTGGCTG 7149

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||  ||  |   |  ||||||||| |||||||||  |||||||    |||| |
Db       7150 GTGCTGGCGGTGATGATCTGCTTCGGTGCGATCGGCTGGTACGACGACTGGATCAAGATC 7209

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |   |   ||  |   ||| ||  |   |||||||  ||   |  |||||||  | |||
Db       7210 GTGCGGCGCGACCCGAACGGCTTGAAGTCGCGCTGGAAATATCTGCTGCAGTCGATCTTC 7269

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||| | |  ||| | | |||||||  |||   ||  ||||     |||  ||  ||  |||| |
Db       7270 GGCCTGGCCGCGGGCCTGTTCCTGTACTACACGGCCGACGTGCCGGCCGCGATCACCTTC 7329

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTG---GGCATCTTCTTCGTGGTCCTG 537
              | | |||||| | |||  | |||  | | | ||   |||  |  |||||||||  |
Db       7330 TACATCCCGATGTTCAAGGCCATCGCCTTGCCGCTGGCCGGCGTGAGCTTCGTGGTGATC 7389

Qy        538 ACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGC 597
              ||||  |||   |||||||||     ||| |||||  |||||  |||||||| || ||||||
Db       7390 GCCTATTTCTGGATCGTCGGCTTCTCCAACGCGGTCAACCTGACCGACGGCCTGGACGGC 7449

Qy        598 CTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCG 657
              |||||||||||||||||  ||  |||  |||||||||||  |||||  |||     |||
Db       7450 CTGGCGATCATGCCGACCGTGCTGGTCGCCTGCGCGCTGGGCGTCTTCGCCTATGCCTCG 7509

Qy        658 GGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTG 717
              ||||||||   ||||||||  |||||   |||  ||      |  || ||||||||  ||||||
Db       7510 GGCAACGTGGTATTCGCCGAATACCTCAAGATCCCGCTGATTCCCGGCGCCGGTGAGCTG 7569

Qy        718 ATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTAT 777
              |||  |  |||||  | |||  |  |||  ||  |||  |  |||||  ||  ||||||||||||||
Db       7570 ATCATCATCTGCTCGGCGATTGCCGGGGCAGGCTTGGGCTTTCTGTGGTTCAACACCTAC 7629

Qy        778 CCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACC 837
              ||||||  |||  ||||||||||||  ||||||||||  ||  |||||||  | |||  ||||||
Db       7630 CCGGCGATGGTGTTCATGGGCGACATCGGCGCGCTGTCGTTGGGCGCGGTGCTCGGCACC 7689

Qy        838 ATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATG 897
              | ||||||||||||||  |||||  | ||||||  | ||||||| ||  ||||||||||
Db       7690 GTGGCGGTGATCGTGCGTCAGGAACTGGTGCTGGTGATCATGGGCGGCGTGTTCGTCATC 7749

Qy        898 GAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTC 957
              |||||  ||  ||||||||||||||||||||||||  ||||||||||||||  |||||  |||
Db       7750 GAAACGCTGTCGGTGATGATCCAGGTCGCCAGCTTCAAGCTGACCGGCAAGCGCGTGTTC 7809

Qy        958 CGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATC 1017
              || |||||||  |||||||||| |||||  ||  || ||||||||||  ||||||||||||
Db       7810 CGCATGGCGCCCATCCATCACCACTTCGAGCTCAAGGGCTGGCCGGAGCCGCGCGTGATC 7869
```

```
Qy       1018 GTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTG 1077
             ||||||||||||||||||| | ||| | |||||||||||||||||| ||||| |||||| ||
Db       7870 GTGCGCTTCTGGATCATCTCGGTGGTGCTGGTGCTGATCGGCCTTGCCACGTTGAAGGTG 7929

Qy       1078 CGTTGA 1083
             || |||
Db       7930 CGCTGA 7935


RESULT 6
AE011708
LOCUS       AE011708               13079 bp    DNA     linear   BCT 29-MAY-2002
DEFINITION  Xanthomonas axonopodis pv. citri str. 306,  section 86 of 469 of
            the complete genome.
ACCESSION   AE011708 AE008923
VERSION     AE011708.1  GI:21106891
KEYWORDS    .
SOURCE      Xanthomonas axonopodis pv. citri str. 306
  ORGANISM  Xanthomonas axonopodis pv. citri str. 306
            Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
            Xanthomonadaceae; Xanthomonas.
REFERENCE   1  (bases 1 to 13079)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
  JOURNAL   Nature 417 (6887), 459-463 (2002)
  MEDLINE   22022145
   PUBMED   12024217
REFERENCE   2  (bases 1 to 13079)
  AUTHORS   da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,
            Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
            Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
            Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
            Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
            Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
            Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
            Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
            Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
            Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
            Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
            Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,
            Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinola,L.A.F.,
            Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
            Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
            Kitajima,J.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-2001) Departmento de Bioquimica, Universidade de
            Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
            Brazil
FEATURES             Location/Qualifiers
     source          1. .13079
                     /organism="Xanthomonas axonopodis pv. citri str. 306"
                     /mol_type="genomic DNA"
                     /strain="306"
                     /db_xref="taxon:190486"
                     /note="pathovar: citri"
     gene            193. .639
                     /gene="XAC0771"
     CDS             193. .639
                     /gene="XAC0771"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAM35659.1"
                     /db_xref="GI:21106892"
                     /translation="MFQGETAITVDDKGRMAVPTAYRDLVARVSGNRLVLTYNPFEAG
                     CLWLYAEKEWERVRDDVMSKPNTQRVVRTLQQKLVGSSAVLELDANGRLSIPASHRNA
                     VGIEKKAVLLGMGDKFELWSEQAHRALIQQTLSDGDLGDELLDLRL"
     gene            684. .1712
                     /gene="mraW"
                     /note="synonym: XAC0772"
     CDS             684. .1712
                     /gene="mraW"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein"
                     /protein_id="AAM35660.1"
                     /db_xref="GI:21106893"
                     /translation="MSQSPAAHVPVLYTQVLDGLQVTENGTYLDGTFGRGGHARGVLE
                     HLGPGGRLLVMDKDPEAIAVAEHSFGGDARVSIHRGSFAGLGQVVAAATVDGILLDLG
                     VSSPQLDVAGRGFSFGKDGPLDMRMDPDSGQSAAEWLAQASDREIADVLWTYGEERQS
                     RRIARAIVARRAEQPLLRTAQLADLIASVMPRGDSKTHPATRSFQAIRIHINRELADL
                     EAGLDAALGALKPGGRLAVISFHSLEDRIVKQFMARYAKAPPSNRRLPEAQPFVATLQ
                     LVSGAIKADDAELHVNPRARSAVLRVAEKLGLGNGESGMGKGNSAAASRFPTADSPFP
                     ASANGDAA"
     gene            1709. .1972
                     /gene="ftsL"
                     /note="synonym: XAC0773"
     CDS             1709. .1972
                     /gene="ftsL"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein"
                     /protein_id="AAM35661.1"
                     /db_xref="GI:21106894"
                     /translation="MSRLLLVVLLACSIASAIGVVYMRHMHRKLFVQLSKLEHTRDEL
                     NIEFGRLQLEQATWAESNRVDQVARARIGMKFPETNDIVVVRP"
     gene            2025. .3893
                     /gene="ftsI"
                     /note="synonym: XAC0774"
     CDS             2025. .3893
                     /gene="ftsI"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx/Glimmer/Genemark"
                     /codon_start=1
                     /transl_table=11
                     /product="penicillin-binding protein 3"
                     /protein_id="AAM35662.1"
                     /db_xref="GI:21106895"
                     /translation="MKAGRNRPRSNFNLRGRLTLVGIALGLCSVTLIGRAAYVQLVNR
                     DFYQRQGEARYLRELPIATSRGMITDRNGEPLAVSTPVESIWVNPQELLRNPDRIPEL
                     AKALGQPLDDLNAKLAQKAGKEFMYLQRRINPDKAHAIVDLKIPGVFSQREFRRFYPQ
                     GEAMAHVLGFTNIDDRGQEGLELAFDEWLRGKPGSKKVVRDARGAIVESVDLVRPAQP
                     GKDLTLSIDRRIQFLAYKELRNALVENNAAGGSMVVMDVATGEVLAMVNLPTYNPNSV
                     TGINPSARRNRAVTDLVEPGSTMKPLTVATALTAGVVTKDTVIDTNPGYMSVGRFTIR
                     DVPRNNGVLNVTGVITRSSNIGAAKIAAKVPDQTFYQSIRNFGYGSAPHSGFPGESAG
                     VVLQPARWSGPSKTTMSYGYGLSVTPLQIAQAYATLGNGGKLTPPTFVRGQHNETRQV
                     ITPEVARQVIDMMETVVTQGGAKGAAILGYHVAGKTGTARKNGANGYERGHYNALFAG
                     LVPATRPRFATVIVINDPQGKVYYGGLVSAPVFHRVMEGALRLMDVPPDDIQSWLAAQ
                     AAGKTGQPVVKPQSVDPDPALVPDPVDEVSAGIPTALAPAPAAAPAQPAPLQESRQ"
     gene            3890. .5377
                     /gene="murE"
                     /note="synonym: XAC0775"
     CDS             3890. .5377
                     /gene="murE"
```

```
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="UDP-N-acetylmuramoylalanyl-D-glutamate-2,
                6-diaminopimelate ligase"
                /protein_id="AAM35663.1"
                /db_xref="GI:21106896"
                /translation="MSRSMALSQLLPDVALTHDVQVSGLVMDSRAVRPGDAFVAIAGF
                GAHGLGFVEQARANGAAAVLFEPPAPDGLPVPADAIAVPGLRARLGVMADQFHGRPSQ
                AMRMVGVTGTNGKTSTVQLLAQALTLLGTPTGTLGTLGVGLYGAAVPTGFTTPLVLPT
                HALLAQLRDDGAQAVAMEVSSHALDQGRVDAVHFDVAVFTNLTRDHLDYHGDMAQYGA
                AKAKLFARAGLKAAVVNLDDAFGHTLFAALDPALRAIGVSSRAHADASVSAQALQLDH
                NGINFALNIRGQVHPVHSPLLGRFNVDNLLAVAGALWAMEFAPAQIASVLGQLQPIHG
                RMNRLGGAHGAPLVVVDYAHTPDALEQALSSLRSHAQDRLICVFGCGGERDTGKRPQM
                AAIAELKADVAIVTDDNPRGEDGDAIVADILRGFARPDAAIVQRDRAAAIHQAIGMAS
                PSDIVLIAGKGHEPYQEVAGVRHAFDDALVAAQALLPRTVLGVRP"
gene            5374. .6885
                /gene="murF"
                /note="synonym: XAC0776"
CDS             5374. .6885
                /gene="murF"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="UDP-N-acetylmuramoylalanyl-D-glutamyl-2,
                6-diaminopimelate-D-alanyl-D-alanyl ligase"
                /protein_id="AAM35664.1"
                /db_xref="GI:21106897"
                /translation="MKLTPLSLIAHWAGGELHGDDAMIDAVGNDTRTLADGSLYVALR
                GERFDGHDFAAEAVARGASGVLVERLLPALSVPQVVVDNTELALARIAAGMQRDRATR
                VLALTGSNGKTSVKALLLSILTEAGRVDGTTVYATPGNRNNEIGLPLAVLAAPDDADV
                AIYEMGAGKPGDIAYLTEIAQPHVALVNNIAPAHLERLGSLLGVAQTKGAIYAALPAG
                GVAVINADDAFGMWFEQQLQSQPVQGRRVLRFGVQASAEITARQLRLHPGGAQFTLVT
                PQGEAQIALPLPGRHNVLNALAATGLALGAGIALPVIAAGLAQARAVAGRQIAQTLPN
                GAVLIDDSYNANPGSLDAAIDALAAAQGEGWLVLGDMRELGAEGVALHAAAGRRARAA
                KLQRLYALGELSAAAAQAFGEGGRVFATHAELIDALQADLVGAASQQQQMQQHASGNT
                QEHAVNQEQQSPSVAALSGAASITILVKGSRGSAMDRVVTALLASAEGASHAA"
gene            6875. .7960
                /gene="mraY"
                /note="synonym: XAC0777"
CDS             6875. .7960
                /gene="mraY"
                /note="identified by sequence similarity; putative; ORF
                located using Blastx/Glimmer/Genemark"
                /codon_start=1
                /transl_table=11
                /product="phospho-N-acetylmuramoyl-pentapeptide-
                transferase"
                /protein_id="AAM35665.1"
                /db_xref="GI:21106898"
                /translation="MLLELARWLQQLESLFGLFNYLTFRGILAALTALFLSLWMGPAV
                IRKLAQFKGGQPIRQDGPQTHFSKAGTPTMGGSLILLTVTLSVLLWGDLRNRYVWLVL
                AVMICFGAIGWYDDWIKIVKRDPNGLKSRWKYLLQSIFGLAAGLFLYYTADVPAAITF
                YIPMFKSIALPLAGVSFVVIAYFWIVGFSNAVNLTDGLDGLAIMPTVLVACALGVFAY
                ASGNVVFAEYLKIPLIPGAGELIIICSAIAGAGLGFLWFNTYPAMVFMGDIGALSLGA
                VLGTIAVIVRQEMVLVIMGGVFVIETLSVMIQVASFKLTGKRVFRMAPIHHHFELKGW
                PEPRVIVRFWIISVVLVLIGLATLKVR"
gene            7960. .9336
                /gene="ftsW"
                /note="synonym: XAC0778"
CDS             7960. .9336
                /gene="ftsW"
                /note="identified by sequence similarity; putative; ORF
```

```
Query Match             54.5%;  Score 590;  DB 1;  Length 13079;
Best Local Similarity   72.1%;  Pred. No. 1.1e-67;
Matches  783;  Conservative    0;  Mismatches  300;  Indels    3;  Gaps    1;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||||| ||   | |||||   | |||||| ||  | | |   |||||  | |||  |
Db       6875 ATGCTGCTTGAGTTGGCCCGTTGGCTGCAGCAACTGGAGAGCCTGTTCGGGCTGTTCAAC 6934

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||| | ||||||||   |   | ||| || |||||||  ||||||| ||||| ||
Db       6935 TATCTGACCTTCCGCGGCATCTTGGCTGCGCTGACGGCGCTGTTCCTGTCGTTGTGGATG 6994

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||    | |||||||  |   |        | ||  |||||||  |  | ||| | |||
Db       6995 GGCCCTGCGGTGATCCGCAAGCTCGCCCAGTTCAAGGGCGGCCAGCCGATCCGCCAGGAC 7054

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || |||||  ||||  | ||  |||   |||||  |||||||||||||||  | ||||||||
Db       7055 GGCCCGCAAACGCATTTTTCCAAGGCCGGCACGCCGACCATGGGCGGTTCGCTGATCCTG 7114

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              | || |  || |          ||||||||||  || ||   ||||||||||||||||  |
Db       7115 CTCACCGTCACCTTGTCGGTGCTGCTGTGGGGCGACCTGCGCAACCGCTACGTGTGGTTG 7174

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||  ||| |  |  ||||||| |||||||||   |||||||     |||| |
Db       7175 GTGCTGGCGGTGATGATCTGCTTCGGTGCGATCGGCTGGTACGACGACTGGATCAAGATC 7234

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              || |    |  ||  |||  ||  |   ||||||||||   |  |||||| | |||
Db       7235 GTCAAACGCGATCCCAATGGTTTGAAGTCGCGCTGGAAGTATCTGCTGCAGTCGATCTTC 7294

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              ||  | | ||| | | |||||||  |||  || |||||     |||   | ||  |||  |
Db       7295 GGACTGGCCGCTGGCTTGTTCCTGTACTACACCGCCGACGTGCCGGCGGCGATCACGTTC 7354

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTG---GGCATCTTCTTCGTGGTCCTG 537
              | | |||||| | |||     ||| | | | | ||  |||  |   |||||||||  |
Db       7355 TACATCCCGATGTTCAAGTCGATCGCGCTGCCGCTGGCCGGCGTGAGCTTCGTGGTGATC 7414

Qy        538 ACCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGC 597
              ||||||||   ||||| ||||    ||| ||||||||||||||||||  || || || |||
Db       7415 GCCTACTTCTGGATCGTGGGCTTCTCCAACGCGGTGAACCTCACCGATGGCCTGGATGGC 7474

Qy        598 CTGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCG 657
              ||||  |||||||||||  ||  ||||  ||| ||||||||||  | |||   ||   |||
Db       7475 TTGGCCATCATGCCGACCGTGCTGGTGGCCTGCGCGCTGGGCGTGTTCGCGTATGCGTCC 7534

Qy        658 GGCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTG 717
              |||||||||  |||||||||  ||||||   |||||     | ||  ||  ||||||||||||
Db       7535 GGCAACGTGGTGTTCGCCGAATACCTGAAAATTCCGCTGATTCCCGGTGCCGGCGAGCTG 7594

Qy        718 ATCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTAT 777
              |||  |  |||||  | |||  | |  ||  || |||  | ||||||| ||||||||||||
Db       7595 ATCATCATCTGCTCGGCGATTGCGGGGGCAGGCTTGGGCTTCCTGTGGTTCAACACCTAC 7654

Qy        778 CCGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACC 837
              |||||    |||  ||||||||||||| ||||||||  || ||||||||||  | |||  ||||||
Db       7655 CCGGCAATGGTGTTCATGGGCGACATCGGCGCGTTGTCGCTGGGCGCAGTGCTCGGCACC 7714

Qy        838 ATCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATG 897
              |||||||||||||||||  |||||| | ||||||| || ||||||||  ||||||| ||
Db       7715 ATCGCGGTGATCGTGCGTCAGGAGATGGTGCTGGTGATCATGGGCGGCGTGTTCGTGATC 7774

Qy        898 GAAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTC 957
              |||||  ||  ||||||||||||||||  ||    ||||||||||||||     ||||| |||
Db       7775 GAAACGCTGTCGGTGATGATCCAGGTGGCCAGCTTCAAGCTGACCGGCAAGCGCGTATTC 7834

Qy        958 CGTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATC 1017
              || |||||||||||||| |||||||||| || || || |||||||| ||||||||||||
Db       7835 CGCATGGCGCCGATCCACCACCATTTCGAGCTCAAGGGTTGGCCGGAGCCGCGCGTGATC 7894

Qy       1018 GTGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTG 1077
              |||||||||||||||||  | ||| | ||||||||||||||| ||||| ||||||  ||
Db       7895 GTGCGCTTCTGGATCATTTCGGTGGTGCTGGTGCTGATCGGCCTTGCCACGTTGAAGGTG 7954

Qy       1078 CGTTGA 1083
              || |||
Db       7955 CGCTGA 7960
```

```
RESULT 7
AE016925/c
LOCUS       AE016925              208524 bp    DNA     linear   BCT 14-OCT-2003
DEFINITION  Chromobacterium violaceum ATCC 12472 section 16 of 16 of the
            complete genome.
ACCESSION   AE016925 AE016825
VERSION     AE016925.1  GI:34105514
KEYWORDS    .
SOURCE      Chromobacterium violaceum ATCC 12472
  ORGANISM  Chromobacterium violaceum ATCC 12472
            Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
            Neisseriaceae; Chromobacterium.
REFERENCE   1  (bases 1 to 208524)
  AUTHORS   .
  CONSRTM   Brazilian National Genome Project Consortium
  TITLE     The complete genome sequence of Chromobacterium violaceum reveals
            remarkable and exploitable bacterial adaptability
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)
  MEDLINE   22882880
   PUBMED   14500782
REFERENCE   2  (bases 1 to 208524)
  AUTHORS   Vasconcelos,A.T.R., de Almeida,D.F., Almeida,F.C., de
            Almeida,L.G.P., de Almeida,R., Goncalves,J.A.A., Andrade,E.M.,
            Antonio,R.V., Araripe,J., de Araujo,M.F.F., Filho,S.A., Azevedo,V.,
            Batista,A.J., Bataus,L.A.M., Batista,J.S., Belo,A., vander Berg,C.,
            Blamey,J., Bogo,M., Bonato,S., Bordignon,J., Brito,C.A.,
            Brocchi,M., Burity,H.A., Camargo,A.A., Cardoso,D.D.P.,
            Carneiro,N.P., Carraro,D.M., Carvalho,C.M.B., Cascardo,J.C.M.,
            Cavada,B.S., Chueire,L.M.O., Pasa,T.B.C., Duran,N., Fagundes,N.,
            Falcao,C.L., Fantinatti,F., Farias,I.P., Felipe,M.S.S.,
            Ferrari,L.P., Ferro,J.A., Ferro,M.I.T., Franco,G.R.,
            Freitas,N.S.A., Furlan,L.R., Gazzinelli,R.T., Gomes,E.A.,
            Goncalves,P.R., Grangeiro,T.B., Grattapaglia,D., Grisard,E.C.,
            Guimaraes,C.T., Hanna,E.S., Hungria,M., Jardim,S.N., Laurino,J.,
            Leoi,L.C.T., Fassarella,L., Lima,A., Loureiro,M.F., Lyra,M.C.P.,
            Macedo,M., Madeira,H.M.F., Manfio,G.P., Maranhao,A.Q.,
            Martins,W.S., di Mauro,S.M.Z., de Medeiros,S.R.B., Meissner,R.V.,
            Menck,C.F.M., Moreira,M.A.M., Nascimento,F.F., Nicolas,M.F.,
            Oliveira,J.G., Oliveira,S.C., Paixao,R.F.C., Parente,J.A.,
            Pedrosa,F.O., Pena,S.J.D., Perreira,J.O., Perreira,M.,
            Pinto,L.S.R.C., Pinto,L.S., Porto,J.I.R., Potrich,D.P.,
            Neto,C.E.R., Reis,A.M.M., Rigo,L.U., Rondinelli,E., dos
            Santos,E.B.P., Santos,F.R., Schneider,M.P.C., Seuanez,H.N.,
            Silva,A.M.R., da Silva,A.L.C., Silva,D.W., Silva,R., Simoes,I.C.,
            Simon,D., Soares,C.M.A., Soares,R.B.A., Souza,E.M., Souza,K.R.L.,
            Souza,R.C., Steffens,M.B.R., Steindel,M., Teixeira,S.R.,
            Urmenyi,T., Vettore,A., Wassem,R., Zaha,A. and Simpson,A.J.G.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de
            Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ
            25651070, Brazil
FEATURES             Location/Qualifiers
     source          1. .208524
                     /organism="Chromobacterium violaceum ATCC 12472"
                     /mol_type="genomic DNA"
                     /strain="ATCC 12472"
                     /db_xref="taxon:243365"
     gene            complement(96. .563)
                     /gene="pilE2"
                     /locus_tag="CV4209"
     CDS             complement(96. .563)
                     /gene="pilE2"
                     /locus_tag="CV4209"
                     /note="identified by sequence similarity; putative; ORF
                     located using GeneMark/Blastx/COG2165"
                     /codon_start=1
                     /transl_table=11
                     /product="type IV pilin"
                     /protein_id="AAQ61869.1"
                     /db_xref="GI:34105515"
                     /translation="MKKQRKQQGFTLIELMIVVAIVGILAAIAIPAYQDYTKRARVSE
                     GLALADAAKTAVTEYYATNNAWASGTAPFNTTYGLAASITGNSVSSINVLANGVVQIA
                     YNSTVSNGALVDLVPTVSAGSVQWYCTYTGSGTTAIASASQLSANWVPSTCRQ"
     gene            838. .2184
                     /gene="bioA"
                     /locus_tag="CV4210"
     CDS             838. .2184
                     /gene="bioA"
                     /locus_tag="CV4210"
                     /EC_number="2.6.1.62"
                     /note="identified by sequence similarity; putative; ORF
                     located using GeneMark/Blastx/COG0161"
                     /codon_start=1
                     /transl_table=11
                     /product="adenosylmethionine-8-amino-7-oxononanoate
                     transaminase"
                     /protein_id="AAQ61870.1"
                     /db_xref="GI:34105516"
                     /translation="MSNQAWLERSFAAVWHPCTQMKRHERLPIVPIASADGVWLTDFD
                     GKRYLDGVSSWWVNLFGHGHPRIKQAVRKQLDSLEHVMLAGFTHRPVVELSERLAALS
                     GLGHAFYGSDGASATEIALKMSFHYWKNVGRPQKTRFVSLENSYHGETVGALAVTDVP
                     LFSSTYAELLKPGLRAPSPDARQAVPGERAADVALRAARSLENLLAKHGGEIAAVIVE
                     PLVQGAAGMAMHDPIYLTELRRLCDQYQVHLIADEIAVGFGRTGSFFAYQQAGITPDF
                     LCLSKGITGGFLPLSCVLTRDEIYQAFYHDDVARGFLHSHSYTGNALACAAALAVLDI
                     FDDEDVLEANRVKAEGFTRQMRPLAERPEIRHFRQRGMIWAFDVDTARSDFALAYFSA
                     MLNRGCLLRPIGKSVYFMPPYTLKSEEMDWLASASVDALDEVLAGESVGGDPTALP"
     gene            2347. .3816
                     /locus_tag="CV4211"
     CDS             2347. .3816
                     /locus_tag="CV4211"
                     /note="identified by sequence similarity; putative; ORF
                     located using Glimmer/Blastx"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAQ61871.1"
                     /db_xref="GI:34105517"
                     /translation="MRQAARPGLVPAGPRTDGSRAALQAAQALSEGAGPLLSAAGERS
                     PARRRGRLAQNAAAGGHQRQAVRQPVPGLQANPARKILPPFAAGARTAQAGAAAVHDD
                     GGAAIPEHQPAKLLSAASRLLAGLPPAVRAGAAARLAGQVAGRRRHAGRHLPPAASAR
                     SDRHQPAVPRRPAIGQATGLRPGATGDAVAAGGPDGGPPGLCAGPGRGCPAALSAGQA
                     IAVPARRLPARTGRRAGRDETQPGTTAQDRGGRRLSGAERRSPAAEPSGRGMAAPAPT
                     QACPGIGRDGGGDHRRRGRHLAPRQRRQLAARRRRPRREQRPPAPAARLLPADCRQPE
                     RIWLPAARPASRAALARRRGAVAGPARPAGRVVSVRSALGIDAALRQGSGMRRRDAGP
                     AAAAGEGHAQHHPQRRQFPTGLAAARQREPAGAAAVARPSLQPVARVQAEGRKRRATG
                     PRFQAASAIPPFPADGIPPQRAILTWRKRRAFSVTLPSL"
     gene            3845. .4963
                     /gene="proB"
                     /locus_tag="CV4212"
     CDS             3845. .4963
                     /gene="proB"
                     /locus_tag="CV4212"
                     /EC_number="2.7.2.11"
                     /note="identified by sequence similarity; putative; ORF
                     located using Glimmer/GeneMark/Blastx/COG0263"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamate 5-kinase"
                     /protein_id="AAQ61872.1"
                     /db_xref="GI:34105518"
                     /translation="MHSVIHASDRIVVKVGSSLVTNDGRGLDLNALARWAEEVAELKR
                     RGKQVVLVSSGAIAEGCQRLGWTVRPKGVHELQAAAAVGQMGLCQAYESAFRSFGLRT
                     AQILLTHEDLADRTRYLNARSTLTSLLNLNVVPIINENDTVATSEIRFGDNDTLGALV
                     TNLIEADALVILTDQRGLYSADPRKHPDAEFIHQAEAGDERLEDMAGGAGSSVGTGGM
                     ITKILAAKRAARSGAATVIASGREPHVLSRLADGEAIGTQLVAATNRMAARKQWLADH
                     LKLAGRLLLDDGAALAIRERGTSLLPVGVSAVEGDFLRGEAVACVDAAGHEVARGLVN
                     YSSDEARQIMRKSTREIEAALGYIVEPELIHRDNMVAL"
     gene            5057. .6130
                     /gene="potF1"
                     /locus_tag="CV4213"
     CDS             5057. .6130
                     /gene="potF1"
                     /locus_tag="CV4213"
                     /note="identified by sequence similarity; putative; ORF
                     located using
```

```
                     Glimmer/GeneMark/Blastx/COG0687/TC:3.A.1.11.2"
                     /codon_start=1
                     /transl_table=11
                     /product="polyamine ABC transporter system,
                     substrate-binding protein"
                     /protein_id="AAQ61873.1"
                     /db_xref="GI:34105519"
                     /translation="MNKIALAALLAVACGSAAAEDVLHIYNWNNALSPDTAKRFEQSC
                     KCKLVQDYYGDNEEMLAKLAAGAKGYDMVFPTAFAVNTLLKQGKLQPLDKSRLPNWKN
                     LNSGYLALNQPFDPGNRYAAPTVVSLTLIGYNATQLKKAGVEGKSNSWALVFDPALLA
                     KIKGKVTVLDSQRELMAAALMYLGKDANSTNPADWKAAAETIRKAKPYWAAFNNQSYI
                     KELTVGNIWVALGYSNDLFQAQQDAKNAKRPFELAYRPQQEGNVLAIDNMTILKDAPR
                     PDLAHKFINFMLDGKNASEISNQIGATNPVKAAEAFFKPQIKANPVIMLDPSKGKYVA
                     LKDLDVKSRRELNRLWSQVKIGR"
     gene            complement(6173. .6757)
                     /locus_tag="CV4214"
     CDS             complement(6173. .6757)
                     /locus_tag="CV4214"
                     /note="identified by sequence similarity; putative; ORF
                     located using Glimmer/GeneMark/Blastx/COG3575"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAQ61874.1"
                     /db_xref="GI:34105520"
                     /translation="MPSEGRLEQQLSGLVLASPWLMRALRAARALGLSDWCIGAGAVR
                     GLVWDALHGFASEPSALSDLDLVYFDAENLSPDREAALQRRLASVHPDLPWEVTNQAA
                     VHLWFADRFGYPVEPLPSLEAALATWPEYATAVGVALNADDSLRIIAPFGLDDLFAMT
                     VRRNPARASLEIYRQRVAQKRYQARWPRVTVVLE"
     gene            complement(7321. .8358)
                     /gene="ruvB"
                     /locus_tag="CV4215"
     CDS             complement(7321. .8358)
                     /gene="ruvB"
                     /locus_tag="CV4215"
                     /note="identified by sequence similarity; putative; ORF
                     located using Blastx"
                     /codon_start=1
                     /transl_table=11
                     /product="holliday junction DNA helicase ruvB"
                     /protein_id="AAQ61875.1"
                     /db_xref="GI:34105521"
                     /translation="MIETDKLFGAAPERRIVTPQRASEQEEALERALRPKLLDEYVGQ
                     KKAREQLEIFIEAAKKRGEALDHVLLFGPPGLGKTTLAHIVAREMGVNLRQTSGPVLE
                     RAGDLAALLTNLEPHDVLFIDEIHRLSPVVEEILYPALEDYQIDIMIGEGPAARSVKI
                     DLPPFTLVGATTRAGMLTNPLRDRFGIVARLEFYNAEELTRIVSRSAGLLNVQLSDDG
                     AFEVAKRSRGTPRIANRLLRRVRDYAEVKSDGVVTMAVADAALAMLDVDPAGLDVMDR
                     KLLQAILEKFSGGPVGLDNVAAAIGESTDTIEDVIEPPLIQQGYLQRTPRGRMATAQA
                     YLHFGLPVKDA"
     gene            8726. .9358
                     /gene="ribB"
                     /locus_tag="CV4216"
     CDS             8726. .9358
                     /gene="ribB"
                     /locus_tag="CV4216"
                     /note="identified by sequence similarity; putative; ORF
                     located using GeneMark/Blastx/COG0108"

  Query Match          54.1%;  Score 585.8;  DB 1;  Length 208524;
  Best Local Similarity  71.9%;  Pred. No. 2.3e-67;
  Matches  780;  Conservative    0;  Mismatches  302;  Indels    3;  Gaps    1;

Qy          2 TGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAGT 61
              ||||  ||   ||||||  ||     ||     |     |  ||  ||||   ||||||  | |
Db     144809 TGCTGCTTTGGCTGGCTGATCTGTTGGGTTCGCATATCCGGGCCTTCAATGTCTTCAATT 144750

Qy         62 ACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTGG 121
              ||   ||  |||||||   |   |   ||  |  |||||||||  |   |  |||||||  ||||
Db     144749 ACACCACGCTGCGCGCGGTGATGGCCGCGTTGACCGCGCTGACCATCTCGCTGTTGCTGG 144690

Qy        122 GGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGACG 181
              | ||  |||  |||||||  |    ||       | |  |  ||||||||||  ||||||||||||||
Db     144689 GCCCGTGGGTGATCCGCAAGCTGACCGAGCTCAAAGTCGGCCAGGCGGTGCGCAACGACG 144630

Qy        182 GTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTTA 241
              ||||||||  |  ||||||    |||  ||||||  ||||||||||||||||  |  ||||||||
Db     144629 GTCCGCAGACCCACCTGGTCAAGGCGGGCACGCCGACCATGGGCGGCTCGCTGATCCTGC 144570

Qy        242 CCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTAG 301
              |||||  ||  |||  ||||||||||||||||||||  ||  ||||||   |||||||||  |
Db     144569 TGGCCATCACCCTCACCACGCTGCTGTGGGCGGACCTGTCCAACAAATACGTGTGGCTGC 144510

Qy        302 TGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTGA 361
              ||||||  ||  |  |||      |||  ||  |  ||||     ||||||  |||||||||
Db     144509 TGCTGGCGGTGATGCTGGGCACCGGCGCGCTGGGCTTTTACGACGACTGGCGCAAGGTGG 144450

Qy        362 TCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTCG 421
              |  | |||  ||  |     |||  ||  |     |  |||       ||||||||||  |||
Db     144449 TGTACAAGGACCCGAAGGGCGTGTCCGCCAAGTTCAAGATGGCGTGGCAGTCGGCCATCG 144390

Qy        422 GCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTGA 481
              ||||||||||| ||||||  ||    |  ||  |||  |     |||       |||   ||||
Db     144389 CCATCGGCGCCGGCGTGTTTCTGATCGCCACCGCCAAGCTGCCGGCGTCCACCGAGCTGA 144330

Qy        482 TCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGG---CATCTTCTTCGTGGTCCTGA 538
              |||||||  |  | ||||  |||||      |  |  ||||   |  ||  ||||   ||  ||||
Db     144329 TCGTGCCCTTCTTCAAGACCGTCGCCTATCCGCTGGGCGCCGTCGGCTTCTGCGTGCTGA 144270

Qy        539 CCTACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCC 598
              ||||  ||||||||||||||  |   |||  |||||  |||||  ||||||||  ||  ||  ||||
Db     144269 CCTATTTCGTCATCGTCGGCACCTCCAACGCGGTCAACCTGACCGACGGCCTGGATGGCC 144210

Qy        599 TGGCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGG 658
              |||||    ||||||||||  ||||  |  |  |  ||||  ||  |||   ||  |  | |
Db     144209 TGGCGGCGTTGCCGACGGTGCTGGTGTCGGCTGGCCTGGCGATTTTCGCGTATGTCGCCG 144150

Qy        659 GCAACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGA 718
              ||  |||  |  ||||  ||  ||||||||      ||||  |  |  ||  ||  ||    ||||  ||
Db     144149 GCCACGCGGTGTTCTCCAAGTACCTGGGCCTGCCCTTCATCCCCGGGGGCGCACGAGGTGG 144090

Qy        719 TCGTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATC 778
              |  ||||||||||  |||  ||  |||||||  ||||  ||||||||  |||||||||  ||||  |
Db     144089 TGGTGTTCTGCGCGGCGATGTGCGGCGCCTGCCTGGGCTTCCTGTGGTTCAACGCCTACC 144030

Qy        779 CGGCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCA 838
              ||||  |||||  ||||||||||||||||||||||||||||||||||||||||||  |||||
Db     144029 CGGCCCAGGTGTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTCGGCACGG 143970

Qy        839 TCGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGG 898
              |||||||||||||||||||||||||||||||||||||||||  |||||  ||  |||||  ||||
Db     143969 TGGCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCCTGATGGGCGGCCTGTTCGTGATGG 143910

Qy        899 AAACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCC 958
              |  | ||  ||||||||||||||||  | ||  ||||||||||||||     |||||  ||||
Db     143909 AGGCGCTGTCGGTGATGATCCAGGTGACCTCGTTCAAGCTGACCGGCAAGCGCGTGTTCC 143850

Qy        959 GTATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCG 1018
              | |||||||||  |  ||||||||||||||  ||||  ||||||     |||  |  |||  | |
Db     143849 GCATGGCGCCGCTGCATCACCATTTCGAATTGAAGGGCTGGAAGGAAACCCAGGTGGTGG 143790

Qy       1019 TGCGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGC 1078
              | ||||||||||||  ||||  ||||  ||||||||||||||||    |  ||  |||||||||
Db     143789 TCCGCTTCTGGATCGTCACCATGATGCTGGTGCTGATCGGCCTGTCGACGCTGAAGCTGC 143730

Qy       1079 GTTGA 1083
              | |||
Db     143729 GCTGA 143725


RESULT 8
AE008699
LOCUS       AE008699               22348 bp    DNA     linear   BCT 23-APR-2003
DEFINITION  Salmonella typhimurium LT2, section 7 of 220 of the complete
            genome.
ACCESSION   AE008699 AE006468
VERSION     AE008699.1  GI:16418608
```

KEYWORDS .
SOURCE Salmonella typhimurium LT2
ORGANISM Salmonella typhimurium LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 22348)
AUTHORS McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W., Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F., Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A., Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W., Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
TITLE Complete genome sequence of Salmonella enterica serovar Typhimurium LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE 2 (bases 1 to 22348)
AUTHORS .
CONSRTM The Salmonella typhimurium Genome Sequencing Project
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/, and Pedro Romero and Peter Karp at EcoCyc; http://ecocyc.PangeaSystems.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

FEATURES Location/Qualifiers

source 1. .22348
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"

gene complement(298. .915)
/gene="leuD"
/note="synonym: STM0110"

CDS complement(298. .903)
/gene="leuD"
/EC_number="4.2.1.33"
/note="3-isopropylmalate dehydratase small subunit. (SW:LEUD_SALTY)"
/codon_start=1
/transl_table=11
/product="3-isopropylmalate isomerase (dehydratase), subunit with LeuC"
/protein_id="AAL19074.1"
/db_xref="GI:16418609"
/translation="MAEKFTQHTGLVVPLDAANVDTDAIIPKQFLQKVTRTGFGAHLF
NDWRFLDEKGQQPNPEFVLNFPEYQGASILLARENFGCGSSREHAPWALTDYGFKVVI
APSFADIFYGNSFNNQLLPVTLSDAQVDELFALVKANPGIKFEVDLEAQVVKAGDKTY
SFKIDDFRRHCMLNGLDSIGLTLQHEDAIAAYENKQPAFMR"

RBS complement(910. .915)
/gene="leuD"
/note="putative RBS for leuD; RegulonDB:STMS1H000490"

gene complement(914. .2327)
/gene="leuC"
/note="synonym: STM0111"

CDS complement(914. .2314)
/gene="leuC"
/EC_number="4.2.1.33"
/note="3-isopropylmalate dehydratase large subunit. (SW:LEU2_SALTY)"
/codon_start=1
/transl_table=11
/product="3-isopropylmalate isomerase (dehydratase), subunit with LeuD"
/protein_id="AAL19075.1"
/db_xref="GI:16418610"
/translation="MAKTLYEKLFDAHVVFEAPNETPLLYIDRHLVHEVTSPQAFDGL
RAHHRPVRQPGKTFATMDHNVSTQTKDINASGEMARIQMQELIKNCNEFGVELYDLNH
PYQGIVHVMGPEQGVTLPGMTIVCGDSHTATHGAFGALAFGIGTSEVEHVLATQTLKQ
GRAKTMKIEVTGNAAPGITAKDIVLAIIGKTGSAGGTGHVVEFCGDAIRALSMEGRMT
LCNMAIEMGAKAGLVAPDETTFNYVKGRLHAPKGRDFDEAVEYWKTLKTDDGATFDTV
VALRAEEIAPQVTWGTNPGQVISVTDIIPDPASFSDPVERASAEKALAYMGLQPGVPL
TDVAIDKVFIGSCTNSRIEDLRAAAEVAKGRKVAPGVQALVVPGSGPVKAQAEAEGLD
KIFIEAGFEWRLPGCSMCLAMNNDRLNPGERCASTSNRNFEGRQGRGGRTHLVSPAMA
AAAAVTGHFADIRSIK"

gene complement(2317. .3422)
/gene="leuB"
/note="synonym: STM0112"

CDS complement(2317. .3408)
/gene="leuB"
/EC_number="1.1.1.85"
/note="3-isopropylmalate dehydrogenase. (SW:LEU3_SALTY)"
/codon_start=1
/transl_table=11
/product="3-isopropylmalate dehydrogenase"
/protein_id="AAL19076.1"
/db_xref="GI:16418611"
/translation="MSKNYHIAVLPGDGIGPEVMAQALKVMDAVRSRFDMRITTSHYD
VGGIAIDNHGHPLPKATVEGCEQADAILFGSVGGPKWENLPPESQPERGALLPLRKHF
KLFSNLRPAKLYQGLEAFCPLRADIAANGFDILCVRELTGGIYFGQPKGREGSGQYEK
AFDTEVYHRFEIERIARIAFESARKRRRKVTSIDKANVLQSSILWREIVNDVAKTYPD
VELAHMYIDNATMQLIKDPSQFDVLLCSNLFGDILSDECAMITGSMGMLPSASLNEQG
FGLYEPAGGSAPDIAGKNIANPIAQILSLALLLRYSLDANDAATAIEQAINRALEEGV
RTGDLARGAAAVSTDEMGDIIARYVAEGV"

RBS complement(2322. .2327)
/gene="leuB"
/note="putative RBS for leuC; RegulonDB:STMS1H000491"

gene complement(3408. .4999)
/gene="leuA"
/note="synonym: STM0113"

CDS complement(3408. .4979)
/gene="leuA"
/EC_number="4.1.3.12"
/note="similar to E. coli 2-isopropylmalate synthase (AAC73185.1); Blastp hit to AAC73185.1 (523 aa), 92% identity in aa 1 - 523"
/codon_start=1
/transl_table=11
/product="2-isopropylmalate synthase"
/protein_id="AAL19077.1"
/db_xref="GI:16418612"
/translation="MSQQVIIFDTTLRDGEQALQASLSAKEKLQIALALERMGVDVME
VGFPVSSPGDFESVQTIARTIKNSRVCALARCVEKDIDVAAQALKVADAFRIHTFIAT
SPMHIATKLRSTLDEVIERAVYMVKRARNYTDDVEFSCEDAGRTPVDDLARVVEAAIN
AGARTINIPDTVGYTMPFEFAGIISGLYERVPNIDKAIISVHTHDDLGIAVGNSLAAV
HAGARQVEGAMNGIGERAGNCALEEVIMAIKVRKDIMNVHTNININHHEIWRTSQTVSQI
CNMPIPANKAIVGSGAFAHSSGIHQDGVLKNRENYEIMTPESIGLNQIQLNLTSRSGR
AAVKHRMEEMGYKDTDYNMDHLYDAFLKLADKKGQVFDYDLEALAFINKQQEEPEHFR
LDYFSVQSGSSDIATASVKLACGEEIKAEAANGNGPVDAIYQAINRITGYDVELVKYD
LNAKGQGKDALGQVDIVVNHHGRRFHGVGLATDIVESSAKAMVHVLNNIWRAAEVEKE
LQRKAQNKENNKETV"

RBS complement(3417. .3422)
/gene="leuB"

```
                     /note="putative RBS for leuB; RegulonDB:STMS1H000492"
     RBS             complement(4994. .4999)
                     /gene="leuA"
                     /note="putative RBS for leuA; RegulonDB:STMS1H000493"
     gene            complement(5067. .5298)
                     /gene="leuL"
                     /note="synonym: STM0114"
     CDS             complement(5067. .5153)
                     /gene="leuL"
                     /note="leu operon leader peptide.  (SW:LPL_SALTY)"
                     /codon_start=1
                     /transl_table=11
                     /product="leu operon leader peptide"
                     /protein_id="AAL19078.1"
                     /db_xref="GI:16418613"
                     /translation="MSHIVRFTGLLLLNAFIVRGRPVGGIQH"
     RBS             complement(5162. .5167)
                     /gene="leuL"
                     /note="putative RBS for leuL; RegulonDB:STMS1H000494"
     -10_signal      complement(5271. .5279)
                     /gene="leuL"
                     /note="putative -10_signal for leuL;
                     RegulonDB:STMLTH004472"
     -35_signal      complement(5290. .5298)
                     /gene="leuL"
                     /note="putative -35_signal for leuL;
                     RegulonDB:STMLTH004472"
     gene            5798. .6754
                     /gene="leuO"
                     /note="synonym: STM0115"
     RBS             5798. .5803
                     /gene="leuO"
                     /note="putative RBS for leuO; RegulonDB:STMS1H000495"
     CDS             5810. .6754
                     /gene="leuO"
                     /note="probable activator protein in leuabcd operon.
                     (SW:LEUO_SALTY)"
                     /codon_start=1
                     /transl_table=11
                     /product="putative LysR family transcriptional regulator"
                     /protein_id="AAL19079.1"
                     /db_xref="GI:16418614"
                     /translation="MPEVKTEKPHLLDMGKPQLRMVDLNLLTVFDAVMQEQNITRAAH
                     TLGMSQPAVSNAVARLKVMFNDELFVRYGRGIQPTARAFQLFGSVRQALQLVQNELPG
                     SGFEPTSSERVFNLCVCSPLDNILTSQIYNRVEKIAPNIHVVFKASLNQNTEHQLRYQ
                     ETEFVISYEEFRRPEFTSVPLFKDEMVLVASRKHPRISGPLLEGDVYNEQHAVVSLDR
                     YASFSQPWYDTPDKQSSVAYQGMALISVLNVVSQTHLVAIAPRWLAEEFAESLDLQIL
                     PLPLKLNSRTCYLSWHEAAGRDKGHQWMEDLLVSVCKR"
     protein_bind    6941. .6953

  Query Match             47.8%;  Score 518.2;  DB 1;  Length 22348;
  Best Local Similarity   67.4%;  Pred. No. 2e-58;
  Matches  730;  Conservative    0;  Mismatches  353;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |   |||||||||  |  ||    | |  ||    |||||   ||||||
Db      17556 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 17615

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||  |||||  | ||||  |||  ||||| || | ||||||||||    | ||  | ||| ||
Db      17616 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 17675

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||  ||  | ||||||  |    |||| |  | |     | |||||||  || ||  |||||
Db      17676 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 17735

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              ||  ||| | ||||||  |    ||    || || ||||| |||||||||  | |||||||
Db      17736 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 17795

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||  | || |  ||    |   ||| | |||||  | |  || ||||  |||||  |||
Db      17796 ACGTCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 17855

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||| ||      |  | |||    |||||  |  || || ||||||||||| |||
Db      17856 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 17915

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |   |||  |  ||   || |||          |||||||  || ||||||  ||||||  ||
Db      17916 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 17975

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |   ||||||  ||| |   || ||  |   |   ||  |  ||  || ||||    |||
Db      17976 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 18035

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||||  |  | ||   ||   |   || |||||  | || | | |  | ||| ||
Db      18036 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 18095

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||  |||||||||||  || | ||  || |||||||| |||||||| || || |||
Db      18096 TACTTTGTCATCGTCGGTACGGGTAACGCCGTGAACCTGACCGACGGGCTTGATGGTCTG 18155

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              |||||  |||||||| ||  | |||||||  ||   |  |   |  |   ||    | | |||
Db      18156 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCCTGGGCGACCGGG 18215

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              |||  ||||  ||||||  |  |||||  ||||  |  |||     ||  |||||||| |
Db      18216 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 18275

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  | |||  | ||||||||  ||  | ||||  | |||||  ||||||||||||
Db      18276 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 18335

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||| || |||||||| |||||  ||||||||  ||||||  |||  ||||||  |||
Db      18336 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGTGCGTTGGGCATTATC 18395

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || |||  |  ||||  |||||  |  |||||  | ||||||| || || || ||  |||||
Db      18396 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 18455

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || |||||  ||  | |||||  | ||||||  || ||    |||||  ||  | ||||||
Db      18456 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 18515

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||| ||||||||||  |  |||||||||||||||||||| ||||||||||||  |||
Db      18516 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 18575

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||||||||||| ||  |  |||| |||||||||| |||||  || ||  |||||  | |||
Db      18576 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGGTACGT 18635

Qy       1081 TGA 1083
              |||
Db      18636 TAA 18638


RESULT 9
AL627265
LOCUS       AL627265               251050 bp    DNA     linear   BCT 04-JUL-2003
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 1/20.
ACCESSION   AL627265 AL513382
VERSION     AL627265.1  GI:16501283
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 251050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
```

```
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 251050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .251050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /mol_type="genomic DNA"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            190. .255
                     /gene="STY0001"
                     /note="synonym: thrL"
     CDS             190. .255
                     /gene="STY0001"
                     /note="Orthologue of E. coli thrL (LPT_ECOLI); Fasta hit
                     to LPT_ECOLI (21 aa), 86% identity in 21 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="thr operon leader peptide"
                     /protein_id="CAD01154.1"
                     /db_xref="GI:16501284"
                     /db_xref="SPTREMBL:Q8XG12"
                     /translation="MNRISTTTITTITITTGNGAG"
     gene            337. .2799
                     /gene="STY0002"
                     /note="synonym: thrA"
     CDS             337. .2799
                     /gene="STY0002"
                     /note="Orthologue of E. coli thrA (AK1H_ECOLI); Fasta hit
                     to AK1H_ECOLI (820 aa), 94% identity in 820 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="aspartokinase I/homoserine dehydrogenase I"
                     /protein_id="CAD01155.1"
                     /db_xref="GI:16501285"
                     /db_xref="GOA:Q8Z9R7"
                     /db_xref="SPTREMBL:Q8Z9R7"
                     /translation="MRVLKFGGTSVANAERFLRVADILESNSRQGQVATVLSAPAKIT
                     NHLVAMIEKTIGGQDALPNISDAERIFSDLLAGLASAQPGFPLARLKMVVEQEFAQIK
                     HVLHGISLLGQCPDSINAALICRGEKMSIAIMAGLLEARGHRVTVIDPVEKLLAVGHY
                     LESTVDIAESTRRIAASQIPADHMILMAGFTAGNEKGELVVLGRNGSDYSAAVLAACL
                     RADCCEIWTDVDGVYTCDPRQVPDARLLKSMSYQEAMELSYFGAKVLHPRTITPIAQF
                     QIPCLIKNTGNPQAPGTLIGASSDDDNLPVKGISNLNNMAMFSVSGPGMKGMIGMAAR
                     VFAAMSRAGISVVLITQSSSEYSISFCVPQSDCARARRAMQDEFYLELKEGLLEPLAV
                     TERLAIISVVGDGMRTLRGISAKFFAALARANINIVAIAQGSSERSISVVVNNDDATT
                     GVRVTHQMLFNTDQVIEVFVIGVGGVGGALLEQLKRQQTWLKNKHIDLRVCGVANSKA
                     LLTNVHGLNLDNWQAELAQANAPFNLGRLIRLVKEYHLLNPVIVDCTSSQAVADQYAD
                     FLREGFHVVTPNKKANTSSMDYYHQLRFAAAQSRRKFLYDTNVGAGLPVIENLQNLLN
                     AGDELQKFSGILSGSLSFIFGKLEEGMSLSQATALAREMGYTEPDPRDDLSGMDVARK
                     LLILARETGRELELSDIVIEPVLPDEFDASGDVTAFMAHLPQLDDAFAARVAKARDEG
                     KVLRYVGNIEEDGVCRVKIAEVDGNDPLFKVKNGENALAFYSHYYQPLPLVLRGYGAG
                     NDVTAAGVFADLLRTLSWKLGV"
     misc_feature    337. .1101
                     /gene="STY0002"
                     /note="Pfam match to entry PF00696 aakinase, Amino acid
                     kinase family, score 207.70, E-value 1.7e-58"
     misc_feature    343. .369
                     /gene="STY0002"
                     /note="PS00324 Aspartokinase signature"
     misc_feature    1819. .2781
                     /gene="STY0002"
                     /note="Pfam match to entry PF00742 Homoserine_dh,
                     Homoserine dehydrogenase, score 535.60, E-value 3.4e-157"
     misc_feature    2314. .2382
                     /gene="STY0002"
                     /note="PS01042 Homoserine dehydrogenase signature"
     gene            2801. .3730
                     /gene="STY0003"
                     /note="synonym: thrB"
     CDS             2801. .3730
                     /gene="STY0003"
                     /note="Orthologue of E. coli thrB (KHSE_ECOLI); Fasta hit
                     to KHSE_ECOLI (310 aa), 94% identity in 308 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="homoserine kinase"
                     /protein_id="CAD01156.1"
                     /db_xref="GI:16501286"
                     /db_xref="GOA:Q8XGP5"
                     /db_xref="SWISS-PROT:Q8XGP5"
                     /translation="MVKVYAPASSANMSVGFDVLGAAVTPVDGTLLGDVVSVEAADHF
                     RLHNLGRFADKLPPEPRENIVYQCWERFCQALGKTIPVAMTLEKNMPIGSGLGSSACS
                     VVAALVAMNEHCGKPLNDTRLLALMGELEGRISGSIHYDNVAPCFLGGMQLMIEENGI
                     ISQQVPGFDEWLWVLAYPGIKVSTAEARAILPAQYRRQDCIAHGRHLAGFIHACYSRQ
                     PQLAAALMKDVIAEPYRARLLPGFSQARQAVSEIGALASGISGSGPTLFALCDKPETA
                     QRVADWLSKHYLQNQEGFVHICRLDTAGARVVG"
     misc_feature    3005. .3196
                     /gene="STY0003"
                     /note="Pfam match to entry PF00288 GHMP_kinases, GHMP
                     kinases putative ATP-binding proteins, score 79.20,
                     E-value 5.7e-21"
     misc_feature    3068. .3103
                     /gene="STY0003"
                     /note="PS00627 GHMP kinases putative ATP-binding domain"
     gene            3734. .5020
                     /gene="STY0004"
                     /note="synonym: thrC"
     CDS             3734. .5020
                     /gene="STY0004"
                     /note="Orthologue of E. coli thrC (THRC_ECOLI); Fasta hit
                     to THRC_ECOLI (428 aa), 93% identity in 428 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="threonine synthase"
                     /protein_id="CAD01157.1"
                     /db_xref="GI:16501287"
                     /db_xref="GOA:Q8Z9R6"
                     /db_xref="SPTREMBL:Q8Z9R6"
                     /translation="MKLYNLKDHNEQVSFAQAVTQGLGKQQGLFFPHDLPEFSLTEID
                     EMLNQDFVSRSAKILSAFIGDEIPQQILEERVRAAFAFPAPVAQVESDVGCLELFHGP
                     TLAFKDFGGRFMAQMLTHISGDKPVTILTATSGDTGAAVAHAFYGLENVRVVILYPRG
                     KISPLQEKLFCTLGGNIETVAIDGDFDACQALVKQAFDDEELKTALGLNSANSINISR
                     LLAQICYYFEAVAQLPQGARNQLVISVPSGNFGDLTAGLLAKSLGLPVKRFIAATNIN
                     DTVPRFLHDGKWAPKATQATLSNAMDVSQPNNWPRVEELFRRKIWRLTELGYAAVDDT
                     TTQQTMRELKAKGYISEPHAAVAYRALRDQLNPGEYGLFLGTAHPAKFKESVESILGE
                     TLALPEALAERADLPLLSHHLPADFAALRKLMMTRQ"
     misc_feature    3962. .4867
                     /gene="STY0004"
                     /note="Pfam match to entry PF00291 PALP,
                     Pyridoxal-phosphate dependent enzyme, score 104.50,
                     E-value 2e-27"
     misc_feature    4022. .4066
                     /gene="STY0004"
                     /note="PS00165 Serine/threonine dehydratases
                     pyridoxal-phosphate attachment site"
     gene            complement(5114. .5887)
```

```
                     /gene="STY0005"
                     /note="synonym: yaaA"
     CDS             complement(5114. .5887)
                     /gene="STY0005"
                     /note="Orthologue of E. coli yaaA (YAAA_ECOLI); Fasta hit
                     to YAAA_ECOLI (258 aa), 86% identity in 257 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="CAD01158.1"
                     /db_xref="GI:16501288"
                     /db_xref="SPTREMBL:Q8Z9R5"
                     /translation="MLILISPAKTLDYQSPLATTRYTQPELLDHSQQLIQQARQLSAP
                     QISRLMGISDKLADLNATRFHDWQPHFTPDNARQAILAFKGDVYTGLQAETFNDADFD
                     FAQQHLRMLSGLYGVLRPLDLMQPYRLEMGIRLENPRGKDLYQFWGDIITDKLNEALE
                     AQGDRVVVNLASEEYFKSVKPKKLDAELIKPVFLDEKNGKFKVVSFYAKKARGLMSRF
                     IIENRLTKPEQLTAFDREGYFFDEETSTQDELVFKRYEQ"
     gene            complement(5966. .7396)
                     /gene="STY0006"
                     /note="synonym: yaaJ"
     CDS             complement(5966. .7396)
                     /gene="STY0006"
                     /note="Similar to Bacillus subtilis amino acid carrier
                     protein alst ALST SW:ALST_BACSU (Q45068; P40743) fasta
                     scores: E(): 0, 45.3% id in 477 aa, and to Alteromonas
                     haloplanktis na(+)-linked D-alanine glycine permease DAGA
                     SW:DAGA_ALTHA (P30144) fasta scores: E(): 0, 36.5% id in
                     315 aa
                     Orthologue of E. coli yaaJ (YAAJ_ECOLI); Fasta hit to
                     YAAJ_ECOLI (476 aa), 76% identity in 476 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative amino-acid transport protein"
                     /protein_id="CAD01159.1"
                     /db_xref="GI:16501289"
                     /db_xref="GOA:Q8Z9R4"
                     /db_xref="SPTREMBL:Q8Z9R4"
                     /translation="MPEFFSFINEILWGSVMIYLLLGAGCWFTWRTGFIQFRYIRQFS
                     RSLKGSLSPQPGGLTSFQALCTSLAARIGSGNLAGVALAIAAGGPGAVFWMWVSAIIG
                     MATSFAECSLAQLYKERDPTGQFRGGPAWYMARGLGMRWMGVVFALFLLVAYGLIFNS
                     VQANAVSRALHFAFNIPPLISGIALAFCALLIIIRGIKGVARLMQWLIPLIALLWVAG
```

Query Match 47.8%; Score 518.2; DB 1; Length 251050;
Best Local Similarity 67.4%; Pred. No. 1.3e-58;
Matches 730; Conservative 0; Mismatches 353; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||  |  ||    | |  ||    |||||  ||||||
Db     147729 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 147788

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| || | ||||||||||    | ||  | ||| ||
Db     147789 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 147848

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | ||||||  |   |||| |  | |    | |||||||  || || |||||
Db     147849 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 147908

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| | |||||| |    ||    || || ||||| |||||||||  | |||||||
Db     147909 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 147968

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || ||| |  ||    |  ||| | ||||  | |  || ||||  ||||| |||
Db     147969 ACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 148028

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||| ||      |  |||    |||| |  || || |||||||||| |||
Db     148029 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 148088

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |  |||  |  ||  || |||      |||||||| || ||||||  ||||||  ||
Db     148089 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 148148

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  ||||||  ||| |    || ||  |    |  ||  | || || |||    |||
Db     148149 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 148208

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||||  |  ||   ||  |   || |||||  | || | | |  ||| ||
Db     148209 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 148268

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||| |||||||||||  ||  | || || || |||||  |||||||  || || || |||
Db     148269 TACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACCGACGGGCTTGATGGTCTG 148328

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||| |||||||| ||  | |||||||  ||   | |   |  |  |   | | ||
Db     148329 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCTTGGGCGACCGGG 148388

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| |||||| | |||||||  |||||  |  |||      || |||||||| | |
Db     148389 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 148448

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  | ||| | |||||||| ||  | || ||| | ||||| ||||||||||
Db     148449 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 148508

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||| || |||||||| |||||  |||||||| |||||  |||| ||||||   |||
Db     148509 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGCGCGTTGGGCATTATC 148568

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| |||||  |  ||||| | |||||||| || || || ||  |||||
Db     148569 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 148628

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||||| ||  | ||||| | |||||| || ||    |||||  ||  | ||||||
Db     148629 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 148688

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||| |||||||||| | ||||||||||||||||||| |||||||||| |||
Db     148689 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 148748

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| ||  | |||| |||||||||| |||||  || ||  |||||  | |||
Db     148749 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGGTACGT 148808

Qy       1081 TGA 1083
              |||
Db     148809 TAA 148811
```

```
RESULT 10
AE016834
LOCUS       AE016834               300169 bp    DNA     linear   BCT 21-MAR-2003
DEFINITION  Salmonella enterica subsp. enterica serovar Typhi Ty2, section 1 of
            16 of the complete genome.
ACCESSION   AE016834 AE014613
VERSION     AE016834.1  GI:29136167
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi Ty2
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi Ty2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Salmonella.
REFERENCE   1  (bases 1 to 300169)
  AUTHORS   Deng,W., Liou,S.R., Plunkett III,G., Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
  TITLE     Comparative Genomics of Salmonella enterica Serovar Typhi Strains
            Ty2 and CT18
  JOURNAL   J. Bacteriol. 185 (7), 2330-2337 (2003)
  MEDLINE   22531367
   PUBMED   12644504
REFERENCE   2  (bases 1 to 300169)
  AUTHORS   Deng,W., Liou,S.-R., Plunkett,G. III, Mayhew,G.F., Rose,D.J.,
            Burland,V., Kodoyianni,V., Schwartz,D.C. and Blattner,F.R.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Laboratory of Genetics, University of
            Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .300169
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi Ty2"
                     /mol_type="genomic DNA"
                     /strain="Ty2"
                     /serovar="Typhi"
                     /sub_species="enterica"
                     /db_xref="taxon:209261"
     gene            190. .255
                     /gene="thrL"
                     /locus_tag="t0001"
     CDS             190. .255
                     /gene="thrL"
                     /locus_tag="t0001"
                     /note="corresponds to STY0001 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="thr operon leader peptide"
                     /protein_id="AAO67735.1"
                     /db_xref="GI:29136168"
                     /translation="MNRISTTTITTITITTGNGAG"
     gene            337. .2799
                     /gene="thrA"
                     /locus_tag="t0002"
     CDS             337. .2799
                     /gene="thrA"
                     /locus_tag="t0002"
                     /note="corresponds to STY0002 from Accession AL513382:
                     Salmonella typhi CT18; homoserine dehydrogenase I"
                     /codon_start=1
                     /transl_table=11
                     /product="aspartokinase I"
                     /protein_id="AAO67736.1"
                     /db_xref="GI:29136169"
                     /translation="MRVLKFGGTSVANAERFLRVADILESNSRQGQVATVLSAPAKIT
                     NHLVAMIEKTIGGQDALPNISDAERIFSDLLAGLASAQPGFPLARLKMVVEQEFAQIK
                     HVLHGISLLGQCPDSINAALICRGEKMSIAIMAGLLEARGHRVTVIDPVEKLLAVGHY
                     LESTVDIAESTRRIAASQIPADHMILMAGFTAGNEKGELVVLGRNGSDYSAAVLAACL
                     RADCCEIWTDVDGVYTCDPRQVPDARLLKSMSYQEAMELSYFGAKVLHPRTITPIAQF
                     QIPCLIKNTGNPQAPGTLIGASSDDDNLPVKGISNLNNMAMFSVSGPGMKGMIGMAAR
                     VFAAMSRAGISVVLITQSSSEYSISFCVPQSDCARARRAMQDEFYLELKEGLLEPLAV
                     TERLAIISVVGDGMRTLRGISAKFFAALARANINIVAIAQGSSERSISVVVNNDDATT
                     GVRVTHQMLFNTDQVIEVFVIGVGGVGGALLEQLKRQQTWLKNKHIDLRVCGVANSKA
                     LLTNVHGLNLDNWQAELAQANAPFNLGRLIRLVKEYHLLNPVIVDCTSSQAVADQYAD
                     FLREGFHVVTPNKKANTSSMDYYHQLRFAAAQSRRKFLYDTNVGAGLPVIENLQNLLN
                     AGDELQKFSGILSGSLSFIFGKLEEGMSLSQATALAREMGYTEPDPRDDLSGMDVARK
                     LLILARETGRELELSDIVIEPVLPDEFDASGDVTAFMAHLPQLDDAFAARVAKARDEG
                     KVLRYVGNIEEDGVCRVKIAEVDGNDPLFKVKNGENALAFYSHYYQPLPLVLRGYGAG
                     NDVTAAGVFADLLRTLSWKLGV"
     gene            2801. .3730
                     /gene="thrB"
                     /locus_tag="t0003"
     CDS             2801. .3730
                     /gene="thrB"
                     /locus_tag="t0003"
                     /note="corresponds to STY0003 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="homoserine kinase"
                     /protein_id="AAO67737.1"
                     /db_xref="GI:29136170"
                     /translation="MVKVYAPASSANMSVGFDVLGAAVTPVDGTLLGDVVSVEAADHF
                     RLHNLGRFADKLPPEPRENIVYQCWERFCQALGKTIPVAMTLEKNMPIGSGLGSSACS
                     VVAALVAMNEHCGKPLNDTRLLALMGELEGRISGSIHYDNVAPCFLGGMQLMIEENGI
                     ISQQVPGFDEWLWVLAYPGIKVSTAEARAILPAQYRRQDCIAHGRHLAGFIHACYSRQ
                     PQLAAALMKDVIAEPYRARLLPGFSQARQAVSEIGALASGISGSGPTLFALCDKPETA
                     QRVADWLSKHYLQNQEGFVHICRLDTAGARVVG"

     gene            3734. .5020
                     /gene="thrC"
                     /locus_tag="t0004"
     CDS             3734. .5020
                     /gene="thrC"
                     /locus_tag="t0004"
                     /note="corresponds to STY0004 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="threonine synthase"
                     /protein_id="AAO67738.1"
                     /db_xref="GI:29136171"
                     /translation="MKLYNLKDHNEQVSFAQAVTQGLGKQQGLFFPHDLPEFSLTEID
                     EMLNQDFVSRSAKILSAFIGDEIPQQILEERVRAAFAFPAPVAQVESDVGCLELFHGP
                     TLAFKDFGGRFMAQMLTHISGDKPVTILTATSGDTGAAVAHAFYGLENVRVVILYPRG
                     KISPLQEKLFCTLGGNIETVAIDGDFDACQALVKQAFDDEELKTALGLNSANSINISR
                     LLAQICYYFEAVAQLPQGARNQLVISVPSGNFGDLTAGLLAKSLGLPVKRFIAATNIN
                     DTVPRFLHDGKWAPKATQATLSNAMDVSQPNNWPRVEELFRRKIWRLTELGYAAVDDT
                     TTQQTMRELKAKGYISEPHAAVAYRALRDQLNPGEYGLFLGTAHPAKFKESVESILGE
                     TLALPEALAERADLPLLSHHLPADFAALRKLMMTRQ"
     gene            complement(5114. .5887)
                     /gene="yaaA"
                     /locus_tag="t0005"
     CDS             complement(5114. .5887)
                     /gene="yaaA"
                     /locus_tag="t0005"
                     /note="corresponds to STY0005 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAO67739.1"
                     /db_xref="GI:29136172"
                     /translation="MLILISPAKTLDYQSPLATTRYTQPELLDHSQQLIQQARQLSAP
                     QISRLMGISDKLADLNATRFHDWQPHFTPDNARQAILAFKGDVYTGLQAETFNDADFD
                     FAQQHLRMLSGLYGVLRPLDLMQPYRLEMGIRLENPRGKDLYQFWGDIITDKLNEALE
                     AQGDRVVVNLASEEYFKSVKPKKLDAELIKPVFLDEKNGKFKVVSFYAKKARGLMSRF
                     IIENRLTKPEQLTAFDREGYFFDEETSTQDELVFKRYEQ"
     gene            complement(5966. .7396)
                     /gene="yaaJ"
                     /locus_tag="t0006"
     CDS             complement(5966. .7396)
                     /gene="yaaJ"
                     /locus_tag="t0006"
                     /note="corresponds to STY0006 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="putative amino-acid transport protein"
                     /protein_id="AAO67740.1"
                     /db_xref="GI:29136173"
                     /translation="MPEFFSFINEILWGSVMIYLLLGAGCWFTWRTGFIQFRYIRQFS
                     RSLKGSLSPQPGGLTSFQALCTSLAARIGSGNLAGVALAIAAGGPGAVFWMWVSAIIG
                     MATSFAECSLAQLYKERDPTGQFRGGPAWYMARGLGMRWMGVVFALFLLVAYGLIFNS
                     VQANAVSRALHFAFNIPPLISGIALAFCALLIIIRGIKGVARLMQWLIPLIALLWVAG
                     SVFICLWHIEQMPGVIASIVKSAFGWQEAAAGAAGYTLTQAITSGFQRGMFSNEAGMG
                     STPNAAAAATSYPPHPVAQGIVQMIGVFSDTIIICTASAMIILLAGNHASHSSTEGIQ
                     LLQHAMVSLTGEWGASFVALIVILFAFSSIVANYIYAENNLFFLRLHNAKAIWLLRLA
                     TLGMVIAGTLISFPLIWQLADMIMACMAITNLTAILLLSPVVYTLAGDYLRQRKLGVR
                     PQFDPRRFPDIEPQLAPDTWDAASRD"
     gene            7665. .8618
                     /gene="talB"
                     /locus_tag="t0007"
     CDS             7665. .8618
                     /gene="talB"
                     /locus_tag="t0007"
                     /note="corresponds to STY0007 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="transaldolase B"
                     /protein_id="AAO67741.1"
```

```
                     /db_xref="GI:29136174"
                     /translation="MTDKLTSLRQFTTVVADTGDIAAMKLYQPQDATTNPSLILNAAQ
                     IPEYRKLIDDAVAWAKQQSSDRAQQVVDATDKLAVNIGLEILKLVPGRISTEVDARLS
                     YDTEASIAKAKRIIKLYNDAGISNDRILIKLASTWQGIRAAEQLEKEGINCNLTLLFS
                     FAQARACAEAGVYLISPFVGRILDWYKANTDKKDYAPAEDPGVVSVTEIYEYYKQHGY
                     ETVVMGASFRNVGEILELAGCDRLTIAPALLKELAESEGAIERKLSFSGEVKARPERI
                     TEAEFLWQHHQDPMAVDKLADGIRKFAVDQEKLEKMIGDLL"
     gene            8729. .9319
                     /gene="mog"
                     /locus_tag="t0008"
     CDS             8729. .9319
                     /gene="mog"
                     /locus_tag="t0008"
                     /note="corresponds to STY0008 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1
                     /transl_table=11
                     /product="molybdopterin biosynthesis Mog protein"
                     /protein_id="AAO67742.1"
                     /db_xref="GI:29136175"
                     /translation="MNTLRIGLVSISDRASSGVYQDKGIPALEEWLASALTTPFELQT
                     RLIPDEQAIIEQTLCELVDEMSCHLVLTTGGTGPARRDVTPDATLAIADREMPGFGEQ
                     MRQISLRFVPTAILSRQVGVIRKQALILNLPGQPKSIKETLEGVKADDGSVSVPGIFA
                     SVPYCIQLLDGPYVETAPEVVAAFRPKSARRENMSS"
     gene            complement(9376. .9942)
                     /gene="yaaH"
                     /locus_tag="t0009"
     CDS             complement(9376. .9942)
                     /gene="yaaH"
                     /locus_tag="t0009"
                     /note="corresponds to STY0009 from Accession AL513382:
                     Salmonella typhi CT18"
                     /codon_start=1

 Query Match             47.8%;  Score 518.2;  DB 1;  Length 300169;
 Best Local Similarity   67.4%;  Pred. No. 1.3e-58;
 Matches  730;  Conservative    0;  Mismatches  353;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||  |  ||   | |  ||    |||||  ||||||
Db     147721 ATGTTAGTTTGGCTGGCCGAGCATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCT 147780

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | ||||  ||| |||||  || | ||||||||||   |||  | ||| ||
Db     147781 TATCTGACGTTTCGCGCCATCGTCAGCCTGTTGACCGCGCTGTTCATCTCTTTATGGATG 147840

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | ||||||  |    |||| |  | |    | |||||||  || || |||||
Db     147841 GGCCCGCGTATGATCGCTCGTCTGCAAAAACTCTCTTTTGGCCAGGTGGTACGTAACGAT 147900

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              || ||| | |||||| |     ||     || || ||||| |||||||||   | |||||||
Db     147901 GGCCCGGAATCGCACTTCAGTAAACGCGGTACGCCGACGATGGGCGGCATCATGATCCTG 147960

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || ||  |  ||   |  ||| | |||||  | |  || ||||  |||||  |||
Db     147961 ACGGCGATTGTGATTTCCGTTCTGTTATGGGCTTACCCGTCTAACCCGTACGTCTGGTGC 148020

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||||||| ||     |    | |||    ||||| |  || || ||||||||||| |||
Db     148021 GTGCTGGTGGTATTGATCGGCTACGGCATTATCGGTTTTGTCGATGACTACCGCAAAGTG 148080

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |   |||  |  ||   || |||      |||||||| || ||||||  ||||||  ||
Db     148081 GTGCGGAAAGATACCAAAGGGCTGATTGCGCGCTGGAAATATTTCTGGATGTCGGTTATC 148140

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |   ||||||  |||  |    || ||  |    |   ||  | ||  || ||||   |||
Db     148141 GCGCTCGGCGTGGCCTTTGCGCTTTATCTCGTCGGGAAAGACACGCCCGCGACCCAACTG 148200

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||||||  |  | ||    ||    |  || |||||  | || | |  |  | ||| ||
Db     148201 GTGGTGCCGTTCTTTAAAGATGTTATGCCGCAATTGGGGCTGTTTTACATTCTGCTGTCC 148260

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||| ||||||||||  || | || || || ||||| ||||||| || || || |||
Db     148261 TACTTTGTCATCGTCGGTACGGGTAACGCCGTAAACCTGACCGACGGGCTTGATGGTCTG 148320

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||||| |||||||| ||  | |||||||  ||   | |   |   |    |  | ||
Db     148321 GCGATTATGCCGACTGTCTTCGTTGCCGCCGGCTTTGCGCTGGTGGCTTGGGCGACCGGG 148380

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||| |||| |||||| | |||||||  |||||  |  |||      || |||||||| |
Db     148381 AACATGAACTTCGCCAATTACCTGCATATTCCGTATTTACGCCATGCGGGCGAGCTGGTG 148440

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  |||| | |||| | |||||||| ||  | || ||| | |||| ||||||||||||
Db     148441 ATTGTCTGTACGGCGATTGTCGGCGCGGGATTAGGATTCTTGTGGTTTAACACCTATCCG 148500

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||| || |||||||| |||||  |||||||| |||||| |||| ||||||   |||
Db     148501 GCGCAGGTTTTTATGGGCGATGTCGGATCGCTGGCGTTGGGCGGCGCGTTGGGCATTATC 148560

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||| |  |||| |||||  | |||||  | ||||||||  || || || ||  |||||
Db     148561 GCCGTGCTGCTGCGTCAGGAATTTCTGCTGGTGATCATGGGCGGCGTCTTTGTGGTGGAA 148620

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||||| ||  | |||||  | |||||  || ||    |||||  ||  | ||||||
Db     148621 ACTCTGTCGGTCATCCTGCAGGTGGGTTCCTTTAAACTACGCGGACAGCGTATTTTCCGT 148680

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||||| |||||||||| |  ||||||||||||||||||| ||||||||||  |||
Db     148681 ATGGCGCCTATCCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTGATTGTG 148740

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              |||||||||| ||  |  |||| |||||||||||| ||||  || ||  |||||  | |||
Db     148741 CGCTTCTGGATTATTTCGCTGATGCTGGTGCTGATTGGCCTGGCAACGCTGAAGGTACGT 148800

Qy       1081 TGA 1083
              |||
Db     148801 TAA 148803


RESULT 11
AX189057
LOCUS       AX189057                1083 bp    DNA     linear   PAT 08-AUG-2001
DEFINITION  Sequence 258 from Patent WO0148209.
ACCESSION   AX189057
VERSION     AX189057.1  GI:15142762
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Forsyth,R.A., Ohlsen,K.L. and Zyskind,J.W.
  TITLE     Genes identified as required for proliferation of E. coli
  JOURNAL   Patent: WO 0148209-A 258 05-JUL-2001;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1083
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
     CDS             1. .1083
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC50738.1"
                     /db_xref="GI:15142763"
                     /db_xref="REMTREMBL:CAC50738"
                     /translation="MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRM
                     IAHLQKLSFGQVVRNDGPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVL
                     VVLVGYGVIGFVDDYRKVVRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQL
```

VVPFFKDVMPQLGLPYILLAYFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWA
TGNMNFASYLHIPYLRHAGELVIVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGA
LGIIAVLLRQEFLLVIMGGVFVVETLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWP
EPRVIVRFWIISLMLVLIGLATLKVR"

ORIGIN

Query Match 46.4%; Score 502.2; DB 6; Length 1083;
Best Local Similarity 66.5%; Pred. No. 4.1e-56;
Matches 720; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              |||| |  |  ||||||||||| |  ||    | |  ||    |||||   ||||||
Db          1 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 60

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | ||||  |||  |||||  |||| |||||||||||  | ||  |||||  ||
Db         61 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 120

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | |||||    |  |||||  |  |       | || ||||  |||||  ||||||
Db        121 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 180

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | ||  |||  |    |||   || ||  |||||||||||||||    .  |||||||
Db        181 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 240

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || ||  ||  |   |||  |   ||||||||||||  |  |  |||||  |  |||||  |||
Db        241 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 300

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||  ||||  ||       |    | |||||  ||  ||||   ||  ||  ||  |||||  |||
Db        301 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 360

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |    ||  || ||   ||  ||      || ||||||||  ||||||  ||||||  |
Db        361 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 420

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |   | ||  ||||  |   ||| |||  |  |  |  ||  |  ||  ||  ||   |||
Db        421 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 480

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | ||  ||  |  | ||   ||    |   ||| ||||  |  ||||  |  |  | ||| |
Db        481 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 540

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||  ||  ||   |||| |||||  |||||  |||||  ||||||||||||||||
Db        541 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 600

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              ||  ||  |||||||| ||| | ||  |||||  |   |  |   |  |   |  | |||
Db        601 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 660

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||  |||| || |||   ||| |||  || ||  |  | |     ||||||  || ||| |
Db        661 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 720

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  |||||  | |||||  ||  ||  ||  ||||||||  |||||  ||||||||||||
Db        721 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 780

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              ||||||||||  ||||||||  ||  ||  ||||||||  |  ||  |  |||  |  ||||  |||
Db        781 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 840

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||  ||   |  |||  |||||   ||  |||||  |  ||  |||||  ||  ||||||||   | |||
Db        841 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 900

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||  ||  ||  ||  ||   |  |||||||   |||||  ||  |||   |||||   ||   |  |||||
Db        901 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 960

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||  |||||  ||||||||  |  ||||||||||||||||||||  ||||||||  || |||
Db        961 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1020

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              || ||||||||  ||  |  ||||  |||||  |||||  ||  ||  ||  ||  |||||  | |||
Db       1021 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1080

Qy       1081 TGA 1083
              | |
Db       1081 TAA 1083
```

RESULT 12
ECMUROY
LOCUS ECMUROY 2608 bp DNA linear BCT 12-SEP-1993
DEFINITION Escherichia coli murD gene and ORF-Y (EC 6.3.2.9).
ACCESSION X51584 X52117
VERSION X51584.1 GI:42058
KEYWORDS membrane protein; murD gene;
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase; unidentified reading frame.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 2608)
AUTHORS Ikeda,M., Wachi,M., Ishino,F. and Matsuhashi,M.
TITLE Nucleotide sequence involving murD and an open reading frame ORF-Y spacing murF and ftsW in Escherichia coli
JOURNAL Nucleic Acids Res. 18 (4), 1058 (1990)
MEDLINE 90192099
PUBMED 2179861
REFERENCE 2 (bases 1 to 2608)
AUTHORS Ikeda,M.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1989) Ikeda M., Institute of Applied Microbiology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113, Japan
REFERENCE 3 (bases 26 to 1354)
AUTHORS Flouret,B.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-1990) Flouret B., URA 1131 du Centre National de la Recherche Scientifique, C N R S, Biochimie Moleculaire et Cellulaire, Batiment 432 Universite Paris-Sud, 91405 Orsay, France
COMMENT Sequence independently determined by [3], but authors accept sequence [1] as definitive.
FEATURES Location/Qualifiers

```
     source          1. .2608
                     /organism="Escherichia coli"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:562"
                     /map="2min"
     RBS             98. .102
                     /note="pot. ribosome binding site"
     CDS             111. .1193
                     /note="unnamed protein product; ORF-Y (AA 1-360)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA35932.1"
                     /db_xref="GI:42059"
                     /db_xref="GOA:P15876"
                     /db_xref="SWISS-PROT:P15876"
                     /translation="MLVWLAEHLVKYYSGFNVFSYLTFRAIVSLLTALFISLWMGPRM
                     IAHLQKLSFGQVVRNDGPESHFSKRGTPTMGGIMILTAIVISVLLWAYPSNPYVWCVL
                     VVLVGYGVIGFVDDYRKVVRKDTKGLIARWKYFWMSVIALGVAFALYLAGKDTPATQL
                     VVPFFKDVMPQLGLFYILLAYFVIVGTGNAVNLTDGLDGLAIMPTVFVAGGFALVAWA
                     TGNMNFASYLHIPYLRHAGELVIVCTAIVGAGLGFLWFNTYPAQVFMGDVGSLALGGA
                     LGIIAVLLRQEFLLVIMGGVFVVETLSVILQVGSFKLRGQRIFRMAPIHHHYELKGWP
                     EPRVIVRFWIISLMLVLIGLATLKVR"
     promoter        1018. .1023
                     /note="pot. -35 region"
```

```
     promoter        1043. .1048
                     /note="pot. -10 region"
     RBS             1181. .1185
                     /note="pot. ribosome binding site"
     CDS             1196. .2512
                     /note="unnamed protein product; murD gene product (AA
                     1-438)"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAA35933.1"
                     /db_xref="GI:42060"
                     /db_xref="GOA:P14900"
                     /db_xref="SWISS-PROT:P14900"
                     /translation="MADYQGKNVVIIGLGLTGLSCVDFFLARGVTPRVMDTRMTPPGL
                     DKLPEAVERHTGSLNDEWLMAADLIVASPGIALAHPSLSAAADAGIEIVGDIELFCRE
                     AQAPIVAITGSNGKSTVTTLVGEMAKAAGVNVGVGGNIGLPALMLLDDECELYVLELS
                     SFQLETTSSLQAVAATILNVTEDHMDRYPFGLQQYRAAKLRIYENAKVCVVNADDALT
                     MPIRGADERCVSFGVNMGDYHLNHQQGETWLRVKGEKVLNVKEMKLSGQHNYTNALAA
                     LALADAAGLPRASSLKALTTFTGLPHRFEVVLEHNGVRWINDSKATNVGSTEAALNGL
                     HVDGTLHLLLGGDGKSADFSPLARYLNGDNVRLYCFGRDGAQLAALRPEVAEQTETME
                     QAMRLLAPRVQPGDMVLLSPACASLDQFKNFEQRGNEFARLAKELG"
ORIGIN

  Query Match             46.4%;  Score 502.2;  DB 1;  Length 2608;
  Best Local Similarity   66.5%;  Pred. No. 3.5e-56;
  Matches  720;  Conservative    0;  Mismatches  363;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |   ||||||||||| |  ||     | |  ||    |||||  ||||||
Db        111 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 170

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              ||  |||||  | |||| |||  |||||  |||| ||||||||||    | |||  ||||| ||
Db        171 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 230

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              ||  ||  | |||||    |   |||||  |  |     |  || ||||  |||||  ||||||
Db        231 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 290

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | ||  |||  |    |||   || ||  ||||||||||||||    |||||||
Db        291 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 350

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              ||  ||  ||  |   |||  |   |||||||||||  | |  |||||  |  |||||  |||
Db        351 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 410

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||  ||||  ||       |  |||||   ||  ||||   ||  ||  ||  |  |||||  |||
Db        411 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 470

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |    ||  ||  ||    ||  ||       ||  ||||||||  ||||||  ||||||  |
Db        471 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 530

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |   | ||  |  ||||  |   ||| |||  |  |  |  ||  ||  ||    |||
Db        531 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 590

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              |  || ||  |  ||    ||   |  ||| ||||  | ||||  |  |  | |||  |
Db        591 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 650

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              ||||||||||| || ||  |  |||| ||||| |||||  ||||| ||||||||||||||||
Db        651 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 710

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              || || |||||||| ||| |  || |||||  |  | |   |   |   |  | | |||
Db        711 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 770

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||  |||| || |||   ||| |||  || ||  |  | |    ||||||  || ||| |

Db        771 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 830

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  ||||  ||||| | ||||| || || || |||||||| ||||| ||||||||||||
Db        831 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 890

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||||||  |||||||| || ||  |||||||  | ||  |  |||  | ||||  |||
Db        891 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 950

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              ||  ||   |  | ||  |||||  || |||||  | ||  |||||  ||  ||||||||  | |||
Db        951 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 1010

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              ||  ||  ||  ||  ||   | |||||||  |||||  ||  |||  |||||  ||  |  |||||
Db       1011 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 1070

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||  |||||  ||||||||  |  ||||||||||||||||||  ||||||||  || |||
Db       1071 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 1130

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||  ||||||||  ||   |  ||||  |||||  |||||  ||  ||  ||  ||  |||||  | |||
Db       1131 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 1190

Qy       1081 TGA 1083
              |||
Db       1191 TAA 1193


RESULT 13
AE000118
LOCUS       AE000118               21757 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 8 of 400 of the complete
            genome.
ACCESSION   AE000118 U00096
VERSION     AE000118.1  GI:1786262
KEYWORDS    .
SOURCE      Escherichia coli K12
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 1 to 21757)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 21757)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 21757)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 21757)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
```

Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES Location/Qualifiers

source 1. .21757
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"

gene complement(77. .163)
/gene="leuL"
/note="synonym: b0075"

CDS complement(77. .163)
/gene="leuL"
/function="leader; Amino acid biosynthesis: Leucine"
/note="f28; 100 pct identical to LPL_ECOLI SW: P09149"
/codon_start=1
/transl_table=11
/product="leu operon leader peptide"
/protein_id="AAC73186.1"
/db_xref="GI:1786263"
/translation="MTHIVRFIGLLLLNASSLRGRRVSGIQH"

promoter complement(197. .226)
/note="factor Sigma70; promoter leu; documented +1 at83735"

protein_bind complement(201. .221)
/note="No predicted promoter"
/bound_moiety="LexA predicted site"

protein_bind complement(201. .221)
/note="central position to predicted promoter: -0.5"
/bound_moiety="LexA predicted site"

protein_bind complement(354. .368)
/note="No predicted promoter"
/bound_moiety="Lrp documented site"

protein_bind complement(354. .368)
/note="central position to leuABCD promoter: -156.5"
/bound_moiety="Lrp documented site"

protein_bind 460. .474
/note="central position to predicted promoter: -28.5"
/bound_moiety="Lrp predicted site"

promoter 462. .490
/note="factor Sigma70; predicted +1 start at 84042"

protein_bind complement(519. .537)
/note="central position to predicted promoter: -319.5"
/bound_moiety="TyrR predicted site"

protein_bind 554. .565
/note="central position to predicted promoter:64"
/bound_moiety="AraC predicted site"

gene 646. .1767
/gene="leuO"
/note="synonym: b0076"

CDS 646. .1767
/gene="leuO"
/function="putative regulator; Amino acid biosynthesis: Leucine"
/note="o373; 100 pct identical to LEUO_ECOLI SW: P10151(290 aa) but contains 59 additional N-ter aa and 24 C-term residues; 100 pct identical to PIR: S40589 but contains 24 additional C-term residues"
/codon_start=1
/transl_table=11
/product="probable transcriptional activator for leuABCD operon"
/protein_id="AAC73187.1"
/db_xref="GI:1786264"
/translation="MTHSTAMDSVFIRTRIFMFSEFYSFCFFLFYMHDKSYSSGLFLC
IPIRERELSVTVELSMPEVQTDHPETAELSKPQLRMVDLNLLTVFDAVMQEQNITRAA
HVLGMSQPAVSNAVARLKVMFNDELFVRYGRGIQPTARAFQLFGSVRQALQLVQNELP
GSGFEPASSERVFHLCVCSPLDSILTSQIYNHIEQIAPNIHVMFKSSLNQNTEHQLRY
QETEFVISYEDFHRPEFTSVPLFKDEMVLVASKNHPTIKGPLLKHDVYNEQHAAVSLD
RFASFSQPWYDTVDKQASIAYQGMAMMSVLSVVSQTHLVAIAPRWLAEEFAESLELQV
LPLPLKQNSRTCYLSWHEAAGRDKGHQWMEEQLVSICKR"

protein_bind 1802. .1816
/note="central position to predicted promoter: -39.5"
/bound_moiety="Lrp predicted site"

promoter 1813. .1839
/note="factor Sigma70; promoter ilvIHp4; documented+1 at 85394"

protein_bind 1833. .1847
/note="central position to predicted promoter: -8.5"
/bound_moiety="Lrp predicted site"

promoter 1838. .1866
/note="factor Sigma70; promoter ilvIHp3; documented+1 at 85420"

protein_bind 1915. .1929
/note="central position to predicted promoter:73.5"
/bound_moiety="Lrp predicted site"

protein_bind 1949. .1963
/note="central position to predicted promoter:107.5"
/bound_moiety="Lrp predicted site"

promoter 1955. .1984
/note="factor Sigma70; promoter ilvIHp2; documented+1 at 85534"

protein_bind 1978. .1992
/note="central position to predicted promoter:136.5"
/bound_moiety="Lrp predicted site"

gene 1995. .3809
/gene="ilvI"
/note="synonym: b0077"

CDS 1995. .3809
/gene="ilvI"
/EC_number="4.1.3.18"
/function="enzyme; Amino acid biosynthesis: Isoleucine, Valine"
/note="o604; 98 pct identical (1 gap) to 522 residues fromILVI_ECOLI SW: P00893 (566 aa)"
/codon_start=1
/transl_table=11
/product="acetolactate synthase III, valine sensitive, large subunit"
/protein_id="AAC73188.1"
/db_xref="GI:1786265"
/translation="MFMSLAANCLSKIGRLMCFTHFFRQTVRQAMEMLSGAEMVVRSL
IDQGVKQVFGYPGGAVLDIYDALHTVGGIDHVLVRHEQAAVHMADGLARATGEVGVVL
VTSGPGATNAITGIATAYMDSIPLVVLSGQVATSLIGYDAFQECDMVGISRPVVKHSF
LVKQTEDIPQVLKKAFWLAASGRPGPVVVDLPKDILNPANKLPYVWPESVSMRSYNPT
TTGHKGQIKRALQTLVAAKKPVVYVGGGAITAGCHQQLKETVEALNLPVVCSLMGLGA
FPATHRQALGMLGMHGTYEANMTMHNADVIFAVGVRFDDRTTNNLAKYCPNATVLHID
IDPTSISKTVTADIPIVGDARQVLEQMLELLSQESAHQPLDEIRDWWQQIEQWRARQC
LKYDTHSEKIKPQAVIETLWRLTKGDAYVTSDVGQHQMFAALYYPFDKPRRWINSGGL
GTMGFGLPAALGVKMALPEETVVCVTGDGSIQMNIQELSTALQYELPVLVVNLNNRYL
GMVKQWQDMIYSGRHSQSYMQSLPDFVRLAEAYGHVGIQISHPHELESKLSEALEQVR
NNRLVFVDVTVDGSEHVYPMQIRGGGMDEMWLSKTERT"

```
     protein_bind    1998. .2012
                     /gene="ilvI"
                     /note="central position to predicted promoter:156.5"
                     /bound_moiety="Lrp predicted site"
     promoter        2017. .2045
                     /gene="ilvI"
                     /note="factor Sigma70; promoter ilvIHp1; documented+1 at
                     85597"
     gene            3812. .4303
                     /gene="ilvH"
                     /note="synonym: b0078"
     CDS             3812. .4303
                     /gene="ilvH"
                     /EC_number="4.1.3.18"

  Query Match             46.4%;  Score 502.2;  DB 1;  Length 21757;
  Best Local Similarity   66.5%;  Pred. No. 2.4e-56;
  Matches  720;  Conservative    0;  Mismatches  363;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||| |  ||    | |  ||    |||||   ||||||
Db      12457 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 12516

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| |||| ||||||||||    | ||  ||||| ||
Db      12517 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 12576

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | |||||   |   ||||| | |     | || ||||  |||| ||||||
Db      12577 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 12636

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | || ||| |    |||   || || ||||||||||||||     |||||||
Db      12637 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 12696

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || || |  |||  |   |||||||||||  | |  ||||| |  ||||| ||
Db      12697 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 12756

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||| |||| ||      |    | |||||  || ||||  || || || || ||||| |||
Db      12757 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 12816

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |    ||  || ||    ||  ||       || |||||||| ||||||  ||||||  |
Db      12817 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 12876

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  | || |  |||| |   ||| ||| |  | | | ||  |  ||  ||  ||   |||
Db      12877 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 12936

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
              | || ||  |  | ||    ||   |  ||| |||| | |||| | |  | |||  |
Db      12937 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 12996

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
              |||||||||||  ||  ||   |||| ||||| ||||| ||||| ||||||||||||||||
Db      12997 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 13056

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
              || || |||||||| ||| | || |||||  |  |  |  |  |     | | |||
Db      13057 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 13116

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
              ||  |||| || |||   ||| ||| || ||  |  | |    |||||| || ||| |
Db      13117 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 13176

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
              |  |||| ||||| | ||||| || || || |||||||| ||||| ||||||||||||
Db      13177 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 13236

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
              |||||||||||| |||||||| || ||  |||||||| | || |  ||| | ||||  |||
Db      13237 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 13296

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
              || ||  |  | || |||||  || |||||  | ||  |||||  || ||||||||  | |||
Db      13297 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 13356

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
              || || ||  ||  ||  | |||||||  |||||  ||  |||  |||||  ||  | |||||
Db      13357 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 13416

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
              |||||  |||||  ||||||||  |  ||||||||||||||||||  ||||||||  ||  |||
Db      13417 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 13476

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
              ||  |||||||| ||  |  ||||  |||||  |||||  ||  ||  ||  ||  |||||  | |||
Db      13477 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 13536

Qy       1081 TGA 1083
              |  |
Db      13537 TAA 13539


RESULT 14
AX191720
LOCUS       AX191720               28277 bp    DNA     linear   PAT 15-AUG-2001
DEFINITION  Sequence 2 from Patent WO0149775.
ACCESSION   AX191720
VERSION     AX191720.1  GI:15209889
KEYWORDS    .
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
  AUTHORS   Iversen,P.L.
  TITLE     Antisense antibacterial cell division composition and method
  JOURNAL   Patent: WO 0149775-A 2 12-JUL-2001;
            Avi Biopharma, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .28277
                     /organism="Escherichia coli"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:562"
ORIGIN

  Query Match             46.4%;  Score 502.2;  DB 6;  Length 28277;
  Best Local Similarity   66.5%;  Pred. No. 2.3e-56;
  Matches  720;  Conservative    0;  Mismatches  363;  Indels    0;  Gaps    0;

Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||| |  ||    | |  ||    |||||   ||||||
Db      12532 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 12591

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  |||||  |||| ||||||||||    | ||  ||||| ||
Db      12592 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 12651

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | |||||   |   ||||| | |     | || ||||  |||| ||||||
Db      12652 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 12711

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | || ||| |    |||   || || ||||||||||||||     |||||||
Db      12712 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 12771

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || || |  |||  |   |||||||||||  | |  ||||| |  ||||| ||
Db      12772 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 12831

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              ||| |||| ||      |    | |||||  || ||||  || || || || ||||| |||
Db      12832 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 12891
```

```
Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
Db      12892 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 12951

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
Db      12952 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 13011

Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db      13012 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 13071

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db      13072 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 13131

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db      13132 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 13191

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db      13192 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 13251

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db      13252 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 13311

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db      13312 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 13371

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db      13372 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 13431

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db      13432 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 13491

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db      13492 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 13551

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Db      13552 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 13611

Qy       1081 TGA 1083
Db      13612 TAA 13614


RESULT 15
EC2MIN
ID   EC2MIN     standard; genomic DNA; PRO; 28277 BP.
XX
AC   X55034; M10429;
XX
SV   X55034.1
XX
DT   21-FEB-1991 (Rel. 27, Created)
DT   05-JUL-1999 (Rel. 60, Last updated, Version 31)
XX
DE   E. coli 2 minute region
XX
KW   ddl gene; envA gene; ftsA gene; ftsQ gene; ftsW gene; ftsZ gene; ilvH gene;
KW   ilvI gene; leuA gene; leuO gene; mraY gene; murC gene; murD gene;
KW   murE gene; murF gene; murG gene; mutT gene; orfA; orfB; orfC; orfX;
KW   pbpB gene; secA gene; shl gene.
XX
OS   Escherichia coli
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Escherichia.
XX
RN   [1]
RP   1-28277
RA   Ayala J.A.;
RT   ;
RL   Submitted (08-JAN-1991) to the EMBL/GenBank/DDBJ databases.
RL   Ayala J.A., Instituto de Biologia Molecular, Centro de Biologia Molecular,
RL   Universidad Autonoma, Canto-Blanco 28049, Madrid, Spain.
XX
RN   [2]
RA   Ayala J.A.;
RT   "Regulation of transcription at the 2-minute region of the genetic map of
RT   Escherichia coli.";
RL   Unpublished.
XX
RN   [3]
RP   1-306
RX   MEDLINE; 82078077.
RX   PUBMED; 6171647.
RA   Wessler S.R., Calvo J.M.;
RT   "Control of leu operon expression in Escherichia coli by a transcription
RT   attenuation mechanism";
RL   J. Mol. Biol. 149(4):579-597(1981).
XX
RN   [4]
RP   268-1130
RX   MEDLINE; 86223773.
RX   PUBMED; 3519576.
RA   Haughn G.W., Wessler S.R., Gemmill R.M., Calvo J.M.;
RT   "High A + T content conserved in DNA sequences upstream of leuABCD in
RT   Escherichia coli and Salmonella typhimurium";
RL   J. Bacteriol. 166(3):1113-1117(1986).
XX
RN   [5]
RP   843-1812
RX   MEDLINE; 88320486.
RX   PUBMED; 3413113.
RA   Henikoff S., Haughn G.W., Calvo J.M., Wallace J.C.;
RT   "A large family of bacterial activator proteins";
RL   Proc. Natl. Acad. Sci. U.S.A. 85(18):6602-6606(1988).
XX
RN   [6]
RP   1799-2187
RX   MEDLINE; 85234358.
RX   PUBMED; 3891724.
RA   Haughn G.W., Squires C.H., DeFelice M., Largo C.T., Calvo J.M.;
RT   "Unusual organization of the ilvIH promoter of Escherichia coli";
RL   J. Bacteriol. 163(1):186-198(1985).
XX
RN   [7]
RP   2101-4431
RX   MEDLINE; 83272971.
RX   PUBMED; 6308579.
RA   Squires C.H., DeFelice M., Devereux J., Calvo J.M.;
RT   "Molecular structure of ilvIH and its evolutionary relationship to ilvG in
RT   Escherichia coli";
RL   Nucleic Acids Res. 11(15):5299-5313(1983).
XX
RN   [8]
RP   4274-6093
RX   MEDLINE; 90330585.
RX   PUBMED; 2198273.
RA   Leclerc G., Noel G., Drapeau G.;
RT   "Molecular cloning, nucleotide sequence and expression of shl, a new gene
RT   in the 2-minute region of the genetic map of Escherichia coli";
RL   J. Bacteriol. 172(8):4696-4700(1990).
XX
RN   [9]
RP   6088-7587
RX   MEDLINE; 90251464.
RA   Gomez M.J., Fluoret B., Van Heijenoort J., Ayala J.A.;
```

RT "Nucleotide sequence of the regulatory region of pbpB gene of Escherichia
RT coli";
RL Nucleic Acids Res. 18:2813-2813(1990).
XX
RN [10]
RP 7316-10074
RX MEDLINE; 83296957.
RX PUBMED; 6350821.
RA Nakamura M., Maruyama I.N., Soma M., Kato J.I., Suzuki H., Hirota Y.;
RT "On the process of cellular division in Escherichia coli: nucleotide
RT sequence of the gene for penicillin-binding protein 3.";
RL Mol. Gen. Genet. 191(1):1-9(1983).
XX
RN [11]
RX MEDLINE; 90124047.
RX PUBMED; 2692800.
RA Tao JS., Ishiguro E.E.;
RT "Nucleotide sequence of the murE gene of Escherichia coli";
RL Can. J. Microbiol. 35(11):1051-1054(1989).
XX
RN [12]
RX MEDLINE; 90328986.
RX PUBMED; 2198024.
RA Michaud C., Parquet C., Flouret B., Blanot D., Van Heijenoort J.;
RT "Revised interpretation of the sequence containing the murE gene encoding
RT the UDP-N-acetylmuramyl-tripeptide synthetase of Escherichia coli";
RL Biochem. J. 269(1):277-278(1990).
XX
RN [13]
RP 11142-12634
RX MEDLINE; 89345095.
RA Parquet C., Fluoret B., Menguin-Lecreulx D., Van Heijenoort J.;
RT "Nucleotide sequence of the murF gene encoding the UDP-MurNac- pentapeptide
RT synthetase of Escherichia coli";
RL Nucleic Acids Res. 17:5379-5379(1989).
XX
RN [14]
RP 12423-15030
RX MEDLINE; 90192099.
RA Ikeda M., Wachi M., Ishino F., Matsuhashi M.;
RT "Nucleotide sequence involving murD and an open reading frame orf-Y spacing
RT murF and ftsW in Escherichia coli";
RL Nucleic Acids Res. 18:1058-1058(1990).
XX
RN [15]
RP 13392-15020
RA Menguin-Lecreulx D, Van Heijenoort J.;
RT "Nucleotide sequence of the murD gene encoding the UDP-MurNac- L-Ala-D-gly
RT synthetase of Escherichia coli";
RL Nucleic Acids Res. 18:183-183(1990).
XX
RN [16]
RP 14743-16239
RX MEDLINE; 90036736.
RX PUBMED; 2509435.
RA Ikeda M., Sato T., Wachi M., Jung H.K., Ishino F., Kobayashi Y.,
RA Matsuhashi M.;
RT "Structural similarity among Escherichia coli FtsW and RodA proteins and
RT Bacillus subtilis SpoVE protein, which function in cell division, cell
RT elongation, and spore formation, respectively";
RL J. Bacteriol. 171(11):6375-6378(1989).
XX
RN [17]
RP 16094-17806
RA Menguin-Lecreulx D., Texier L., Van Heijenoort J.;
RT "Nucleotide sequence of the cell-envelope murG gene of Escherichia coli";
RL Nucleic Acids Res. 18:2810-2810(1990).
XX
RN [18]
RP 16094-18886
RX MEDLINE; 90326550.
RA Ikeda M., Wachi M., Jung H.K., Ishino F., Matsuhashi M.;
RT "Nucleotide sequence involving murG and murC in the mra gene cluster region
RT of Escherichia coli";
RL Nucleic Acids Res. 18:4014-4014(1990).
XX
RN [19]
RP 19464-21952
RX MEDLINE; 85054557.
RX PUBMED; 6094474.
RA Robinson A.C., Kenan D.J., Hatfull G.F., Sullivan N.F., Spiegelberg R.,
RA Donachie W.D.;
RT "DNA sequence and transcriptional organization of essential cell division
RT genes ftsQ and ftsA of Escherichia coli: evidence for overlapping
RT transcriptional units.";
RL J. Bacteriol. 160(2):546-555(1984).
XX
RN [20]
RP 21464-23333
RX MEDLINE; 86083166.
RX PUBMED; 3000876.
RA Yi Q.M., Lutkenhaus J.;
RT "The nucleotide sequence of the essential cell-division gene ftsZ of
RT Escherichia coli";
RL Gene 36(3):241-247(1985).
XX
RN [21]
RP 22964-25011
RX MEDLINE; 88058745.
RX PUBMED; 2824434.
RA Beall B., Lutkenhaus J.;
RT "Sequence analysis, transcriptional organization, and insertional
RT mutagenesis of the envA gene of Escherichia coli";
RL J. Bacteriol. 169(12):5408-5415(1987).
XX
RN [22]
RP 23989-27799
RX MEDLINE; 88298644.
RX PUBMED; 2841285.
RA Schmidt M., Rollo E., Grodberg I., Oliver D.;

Query Match 46.4%; Score 502.2; DB 15; Length 28277;
Best Local Similarity 66.5%; Pred. No. 2.3e-56;
Matches 720; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```
Qy          1 ATGCTCCTGCTGCTGGCCGAATACCTGCAACAGTTCTACAAGGGCTTCGGCGTCTTCCAG 60
              ||| |  |  |||||||||| |  ||    | |  ||    |||||  ||||||
Db      12532 ATGTTAGTTTGGCTGGCCGAACATTTGGTCAAATATTATTCCGGCTTTAACGTCTTTTCC 12591

Qy         61 TACCTGACCCTGCGCGGCATTCTCAGCGTGCTCACCGCGCTGTCGCTGTCGCTGTGGCTG 120
              || |||||  | |||| |||  ||||| |||| ||||||||||    | ||  ||||| ||
Db      12592 TATCTGACGTTTCGCGCCATCGTCAGCCTGCTGACCGCGCTGTTCATCTCATTGTGGATG 12651

Qy        121 GGGCCCTGGATGATCCGTACCTTGCAGATCCCCCAGATCGGCCAGGCCGTGCGCAACGAC 180
              || ||  | |||||   |  |||||  |  |      | || ||||  ||||| ||||||
Db      12652 GGCCCGCGTATGATTGCTCATTTGCAAAAACTTTCCTTTGGTCAGGTGGTGCGTAACGAC 12711

Qy        181 GGTCCGCAGTCGCACCTGTCGAAGAAGGGCACCCCGACCATGGGCGGCGCCCTGATCCTT 240
              |||||  | || ||| |   |||   || || ||||||||||||||     |||||||
Db      12712 GGTCCTGAATCACACTTCAGCAAGCGCGGTACGCCGACCATGGGCGGGATTATGATCCTG 12771

Qy        241 ACCGCCATAGCCATCAGCACGCTGCTGTGGGCGGATCTTTCCAACCGCTACGTGTGGGTA 300
              || || ||  |  |||  |  |||||||||||  |  |  |||||  |  |||||  |||
Db      12772 ACGGCGATTGTGATCTCCGTACTGCTGTGGGCTTACCCGTCCAATCCGTACGTCTGGTGC 12831

Qy        301 GTGCTGGTCGTTACCCTGCTGTTCGGTGCCATCGGCTGGGTAGACGACTACCGCAAGGTG 360
              |||  |||| ||      |   | |||||  || ||||  || || || || || ||||| |||
Db      12832 GTGTTGGTGGTGCTGGTAGGTTACGGTGTTATTGGCTTTGTTGATGATTATCGCAAAGTG 12891

Qy        361 ATCGAGAAGAACTCCCGTGGCCTGCCGAGCCGCTGGAAGTACTTCTGGCAGTCGGTGTTC 420
              |    ||  || ||   ||  ||       || |||||||| ||||||  ||||||  |
Db      12892 GTGCGTAAAGACACCAAAGGGTTGATCGCTCGTTGGAAGTATTTCTGGATGTCGGTCATT 12951

Qy        421 GGCATCGGCGCCGCCGTGTTCCTCTACATGACTGCCGAAACCCCGATCGAGACCACCCTG 480
              |  | ||  | ||||  |   ||| |||  |  | | | |  | ||  ||  ||    |||
Db      12952 GCGCTGGGTGTCGCCTTCGCCCTGTACCTTGCCGGCAAAGACACGCCCGCAACGCAGCTG 13011
```

```
Qy        481 ATCGTGCCGATGCTGAAGAGCGTCGAGATCCAGTTGGGCATCTTCTTCGTGGTCCTGACC 540
Db      13012 GTGGTCCCATTCTTTAAAGATGTGATGCCGCAGCTGGGGCTGTTCTACATTCTGCTGGCT 13071

Qy        541 TACTTCGTCATCGTCGGCTCGAGCAATGCGGTGAACCTCACCGACGGTCTCGACGGCCTG 600
Db      13072 TACTTCGTCATTGTGGGTACTGGCAACGCGGTAAACCTGACCGATGGTCTCGACGGCCTG 13131

Qy        601 GCGATCATGCCGACGGTAATGGTTGCCGGCGCGCTGGGCATCTTCTGCTACCTGTCGGGC 660
Db      13132 GCAATTATGCCGACCGTATTTGTCGCCGGTGGTTTTGCGCTGGTGGCGTGGGCGACCGGC 13191

Qy        661 AACGTGAAGTTCGCCGAGTACCTGCTGATTCCCAACGTACCGGGCGCCGGCGAGCTGATC 720
Db      13192 AATATGAACTTTGCCAGCTACTTGCATATACCGTATCTGCGACACGCCGGGGAACTGGTT 13251

Qy        721 GTGTTCTGCGCCGCGCTGGTCGGCGCCGGCCTCGGCTTCCTCTGGTTCAACACCTATCCG 780
Db      13252 ATTGTCTGTACCGCGATAGTCGGGGCAGGACTGGGCTTCCTGTGGTTTAACACCTATCCG 13311

Qy        781 GCGCAGGTCTTCATGGGCGACGTCGGCGCGCTGGCGCTGGGCGCCGCGCTGGGCACCATC 840
Db      13312 GCGCAGGTCTTTATGGGCGATGTAGGTTCGCTGGCGTTAGGTGGTGCGTTAGGCATTATC 13371

Qy        841 GCGGTGATCGTGCGCCAGGAGATCGTGCTGTTCATCATGGGTGGGGTGTTCGTCATGGAA 900
Db      13372 GCCGTACTGCTACGTCAGGAATTCCTGCTGGTGATTATGGGGGGCGTGTTCGTGGTAGAA 13431

Qy        901 ACCCTCTCGGTGATGATCCAGGTCGCTTCCTTCAAGCTGACCGGACGCCGCGTCTTCCGT 960
Db      13432 ACGCTTTCTGTCATCCTGCAGGTCGGCTCCTTTAAACTGCGCGGACAACGTATTTTCCGC 13491

Qy        961 ATGGCGCCGATCCATCACCATTTCGAACTGAAAGGCTGGCCGGACCCGCGCGTGATCGTG 1020
Db      13492 ATGGCACCGATTCATCACCACTATGAACTGAAAGGCTGGCCGGAACCGCGCGTCATTGTG 13551

Qy       1021 CGCTTCTGGATCATCACCGTGATCCTGGTGCTGATCGGCCTCGCCACCTTGAAGCTGCGT 1080
Db      13552 CGTTTCTGGATTATTTCGCTGATGCTGGTTCTGATTGGTCTGGCAACGCTGAAGGTACGT 13611

Qy       1081 TGA 1083
Db      13612 TAA 13614

Search completed: May  9, 2004, 14:51:52
Job time : 4400 secs
```